OUK-H05969-16-KP2364

THE JOURNAL OF PHILOSOPHY

VOLUME CI, NUMBER 1
JANUARY 2004

16

page
5 *In Search of a Substantive Theory of Truth* Gila Sher
COMMENTS AND CRITICISM
37 *Hardin, Tye, and Color Physicalism* Alex Byrne and David R. Hilbert
BOOK REVIEWS
44 A.D. Smith: *The Problem of Perception* Maura Tumulty
54 MEMO TO AUTHORS

Published by **The Journal of Philosophy, Inc.**

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Akeel Bilgrami, Arthur C. Danto, Kent Greenawalt, Patricia Kitcher, Philip Kitcher, Isaac Levi, Mary Mothersill, Philip Pettit, Carol Rovane, Achille C. Varzi. *Editor Emeritus*: Sidney Morgenbesser. *Consulting Editors*: David Albert, John Collins, James T. Higginbotham, Charles D. Parsons, Wilfried Sieg. *Managing Editor*: John Smylie.

THE JOURNAL OF PHILOSOPHY is owned and published by the Journal of Philosophy, Inc. *President*, Arthur C. Danto; *Vice President*, Akeel Bilgrami; *Secretary*, Daniel Shapiro; *Treasurer*, Barbara Gimbel; *Other Trustees*: Lee Bollinger, Leigh S. Cauman, Kent Greenawalt, Michael J. Mooney, Lynn Nesbit.

All communications to the Editors and Trustees and all manuscripts may be sent to John Smylie, Managing Editor, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, New York 10027. FAX: (212) 932-3721.
You may also visit our website at: www.journalofphilosophy.org

THE JOURNAL OF PHILOSOPHY

2004

SUBSCRIPTIONS (12 issues)

Individuals	$35.00
Libraries and Institutions	$75.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$15.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are required.

Published monthly as of January 1977; typeset and printed by Capital City Press, Montpelier, VT.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL's terms for all copying beyond that permitted by Sections 107 or 108 of the U S Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

POSTMASTER: Periodical postage paid at New York, NY, and other mailing offices
POSTMASTER: Send address changes to the *Journal of Philosophy* at MC 4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027

© Copyright 2004 by the Journal of Philosophy, Inc. ISSN 0022-362X

THE JOURNAL OF PHILOSOPHY was founded in 1904 as THE JOURNAL OF PHILOSOPHY, PSYCHOLOGY AND SCIENTIFIC METHODS by Frederick J. E. Woodbridge and J. McKeen Cattell. In 1906, Wendell T. Bush became associated with the JOURNAL as co-editor. The title was changed in 1921, and the JOURNAL was incorporated in the State of New York under its present name in 1923. Since its founding, it has been published from the Department of Philosophy at Columbia University. It is typeset and printed by Capital City Press.

THE JOURNAL OF
PHILOSOPHY

VOLUME CI, NUMBER 1
JANUARY–DECEMBER 2004

Edited by
Bernard Berofsky, Akeel Bilgrami
Arthur C. Danto, Kent Greenawalt
Patricia Kitcher, Philip Kitcher
Isaac Levi, Mary Mothersill, Philip Pettit
Carol Rovane, Achille C. Varzi

Consulting Editors
David Albert, John Collins
James T. Higginbotham, Charles D. Parsons
Wilfried Sieg

Managing Editor
John Smylie

Published by The Journal of Philosophy, Inc.

KP 2364

100
J 826

STATEMENT OF COPYRIGHT POLICY

We would like to draw your attention to the protection copyright affords against multiple photocopying by libraries, instructors, and copy centers. The 1976 copyright law provides that photocopies in excess of "fair use" copying cannot be made without authorization by the copyright owner—"fair use" meaning single copying for personal use. Accordingly, the JOURNAL authorizes personal or educational multiple copying only if a request is made in writing, and the fee of $.05 per page/per copy is paid directly to us. *Each copy must include a notice of copyright.* We shall turn over to the authors any substantial payments (over $30.00, an administrative fee) obtained from all photocopying. Inquiries about photocopying may be sent to the Managing Editor.

© Copyright 2004 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER Second class postage paid at New York, NY, and other mailing offices. Please send address changes to THE JOURNAL OF PHILOSOPHY, 709 Philosophy Hall, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027. Fax (212) 932-3721

THE JOURNAL OF PHILOSOPHY
VOLUME CI, NO. 1, JANUARY 2004

IN SEARCH OF A SUBSTANTIVE THEORY OF TRUTH*

> The layman...expects philosophers to answer deep questions of great import for an understanding of the world.... And the layman is quite right.... Yet he finds most writings by philosophers of the analytical school disconcertingly remote from these concerns.... The complaint...is understandable.... [A]nalytical philosophy passed, comparatively recently, through a destructive phase.... During that phase, it appeared as though demolition was the principal legitimate task of philosophy. Now most of us believe once more that philosophy has a constructive task; but, so thoroughly was the demolition accomplished, that the rebuilding is of necessity slow. —Michael Dummett[1]

Is a substantive theory of truth feasible? What would be the scope, structure, and content of such a theory? My idea of "a substantive theory" has the everyday connotation of "a theory that provides an explanatory, constructive, and systematic account of a rich, significant, and fundamental subject-matter." "A substantive theory of truth" in this sense contrasts with "a deflationist theory of truth." Where deflationists say that "truth is entirely captured by the...triviality...that each proposition specifies its own condition for being true,"[2] advocates of a substantive theory of truth (henceforth, *substantivists*) say that truth is not entirely captured by this triviality; where deflationists say that "the truth predicate exists solely for the sake of a certain logical need"[3] (that is, indirect reference to, and generalization over, propositions), substantivists say that it exists for other needs as well; where deflationists say that truth is not a deep notion, substantivists say it is; and

---

* I would like to thank Roberta Ballarin, Yemima Ben Menachem, Charles Parsons, Carl Posy, Oron Shagrir, Peter Sher, Mark Steiner, Cory Wright, and the participants in my recent graduate seminar on truth for very helpful comments and conversation.

[1] *The Logical Basis of Metaphysics* (Cambridge: Harvard, 1991), p. 1.

[2] Paul Horwich, *Truth* (Cambridge: Blackwell, 1990), p. xi.

[3] Horwich, p. 2. (A compilation of two sentences in inverted order.)

0022-362X/04/0101/5–36

where deflationists say that a theory of truth cannot be, or need not be, genuinely explanatory, substantivists say it can and should be. Substantivists accept the deflationist claim that truth is not mysterious, but they believe truth yields itself to substantive inquiry. My notion of a substantive theory of truth is close to what some call an "inflationary theory,"[4] but "substantive," in its everyday usage, better captures the ordinary, common-sense considerations that motivate me. Like Michael Dummett, I believe that philosophy should take on important "questions...for...understanding...the world," and, like the early Hartry Field, I think that "we'd be crazy to give...up in [philosophy]...a methodology [of substantive theorizing] that has proved extremely fruitful in science."[5]

The attempts to construct a substantive theory of truth, however, have come upon great difficulties, and many philosophers have given up hope of ever producing such a theory. Field, for example, has renounced his plan for a substantive theory of truth (based on a causal account of reference), and today he, along with many adherents of his original plan, is an avid champion of deflationism. Not all contemporary philosophers, however, are satisfied with the prevalent trend. Donald Davidson, Michael Devitt, Anil Gupta, Michael Lynch, Hilary Putnam, Crispin Wright, and others (including me in an earlier article) have dissented, to a greater or lesser extent, from mainstream deflationism.[6]

---

[4] See, for example, Hartry Field, "Deflationist Views of Meaning and Content," *Mind*, CIII (1994): 249–84.

[5] Field, "Tarski's Theory of Truth," this JOURNAL, LXIX, 13 (July 13, 1972): 347–75, here p. 363. The full sentence is: "This is a methodology that has proved extremely fruitful in science, and I think we'd be crazy to give it up in linguistics." Field treats the theory of truth as part of linguistics, but what he says is directed to whatever discipline the theory of truth belongs to, that is, on our demarcation, philosophy. He specifically refers to reductionist theories which, he says, are pointless unless substantive. I think it is reasonable to presume that he extends this point to scientific theories in general. Note: some philosophers identify "substantive theory of truth" with "correspondence theory of truth"; I prefer to distinguish between the two. Although I will eventually advocate a substantive correspondence theory of truth, I do not want to rule out in advance the possibility of either a substantive noncorrespondence theory of truth or a nonsubstantive correspondence theory.

[6] See Davidson, "The Structure and Content of Truth," this JOURNAL, LXXXVII, 6 (June 1990): 279–328, and "The Folly of Trying to Define Truth," this JOURNAL, XCIII, 6 (June 1996): 263–78; Devitt, *Realism and Truth*, 1st/2nd edition (Cambridge: Blackwell, 1984/91); Gupta, "Minimalism," *Philosophical Perspectives*, VII (1993): 359–69, and "A Critique of Deflationism," *Philosophical Topics*, XXI, 2 (1993): 57–81; Lynch, *Truth in Context* (Cambridge: MIT, 1998); Putnam, *Reason, Truth, and History* (New York: Cambridge, 1981), and *The Many Faces of Realism* (La Salle, IL: Open Court, 1987); Wright, *Truth and Objectivity* (Cambridge: Harvard, 1992), and "Truth: A Traditional Debate Reviewed," in Simon Blackburn and Keith Simmons, eds., *Truth* (New York: Oxford, 1999), pp. 203–38; and Gila Sher, "On the Possibility of a Substantive Theory of Truth," *Synthese*, CXVII (1998–99): 133–72.

Here I will further pursue the attempt to construct a substantive theory of truth by investigating some of the challenges facing it and offering a few ideas about how to meet them. In particular, I will connect the methodological challenges facing the theorist of truth with those facing the natural (and social) scientist. This will enable me to place the debate on truth in a new, broader perspective, and point to new ways of approaching the issues. I will concentrate on two complementary challenges: the challenge of *disunity*, and the challenge of *unity*. In the case of truth, these are the challenges of (i) recognizing the diversity, complexity and multidimensionality of truth, and (ii) uncovering its unifying principles.

This article is divided into two sections: (I) Disunity and (II) Unity. Section I presents two disunity challenges: a radical challenge, conducive to deflationism, and a moderate challenge, compatible with a substantive theory. I argue that the radical challenge is unsound but the moderate challenge is a genuine challenge, confronting any theory of a broad and diverse subject matter. To meet this challenge, I propose a few commonsensical guidelines, and I note a few similarities and differences between my approach and that of earlier philosophers (specifically Kant, James, and Wittgenstein).

One ramification of the disunity challenge is moderate pluralism. Moderate pluralism with respect to truth has recently been advanced by Wright and Lynch. Wright, for example, raises the possibility that truth in physics is based on correspondence while truth in mathematics is based on coherence. My own analysis suggests a different kind of pluralism: pluralism within the bounds of correspondence. The idea is that truth both in physics and in mathematics is based on correspondence, but since physics and mathematics involve different aspects of language and the world, their correspondence principles differ. This kind of pluralism brings us closer to the ideal of a balanced theory: a theory balancing the demands of unity with those of diversity. Two problems for any pluralist conception of truth are: (i) In what sense are diverse principles of truth principles of the same thing, namely, truth? and (ii) How can logical inference transmit truth from sentences governed by one type of truth to sentences governed by another? Solutions to both problems are offered in section II.

Section II: The debate on unity and diversity in science sometimes gives the impression that recognition of diversity and a search for unity are incompatible. Like many philosophers, I believe that unity and diversity complement rather than exclude each other, and neither has priority over the other. Wright, too, accepts this view, but his conception of the unifying principles differs from mine. While Wright's conception is minimalistic (the unifying principles are mere "platitudes"), I believe the unifying principles can and should be substantive. Unity is linked to substantiveness through its role in explanation, and the question is not whether substantive unification is possible, but what kind(s) of substantive unifiers are available in our field. I point out two types of such unifiers, *core unifier* and *specialized unifier,* and I formulate two theses exemplifying these types of unifier, the *Immanence Thesis* and the *Logicality Thesis.* The first thesis has implications for correspondence and skepticism; the second yields a new solution to the problem of logical inference across types of truth.

I. THE DISUNITY OF TRUTH[7]

I.A. Radical Disunity Challenge. A well-known argument against the feasibility of a substantive theory of truth says that since every thought (proposition, belief, cognition, judgment, sentence of a given language) has its own unique truth condition, a general and substantive account of truth is impossible. I will call this argument "the radical disunity argument." One formulation of this argument is:

> [C]ompare 'is true'...with a genuine target of philosophical analysis.... We know *individually* what makes ['is true'] applicable to the judgements or sentences of an understood language. 'Penguins waddle' is a sentence true, in English, if and only if penguins waddle. It is true that snow is white if and only if snow is white. The reason the first sentence deserves the predicate is that penguins waddle, and the reason why the judgement that snow is white deserves the predicate is that snow *is* white. But these reasons are entirely different. There is no single account, or even little family of accounts, in virtue of which each deserves the predicate, for deciding whether penguins waddle has nothing much in common with deciding whether snow is white. There are *as* many different things to do, to decide whether the predicate applies, as there are judgements to make. So how *can* there be a unified, common account of the "property" which these quite different decision procedures supposedly determine?[8]

The theoretical principle underlying the radical disunity argument is clearly expressed by Kant in the introduction to "Transcendental Logic" of the *Critique of Pure Reason.* Inquiring whether it is possible to go beyond a minimalist characterization (literally, name clarification) of truth as agreement of a cognition with its object and provide a criterion that determines the precise conditions under which each cognition is true, Kant reasons:

---

[7] This part of the article continues my earlier discussion of the disunity of truth in "On the Possibility of a Substantive Theory of Truth." It is, however, self-sufficient.

[8] Simon Blackburn, *Spreading the Word* (New York: Oxford, 1984), p. 230. This argument may also be interpreted as an argument against the existence of a *property* of truth, but here I am interested in it as an argument against the possibility of a substantive account of truth. Blackburn, it should be noted, does not endorse this argument.

> If truth consists in the agreement of a cognition with its object, then this object must thereby be distinguished from others; for a cognition is false if it does not agree with the object to which it is related even if it contains something that could well be valid of other objects. Now a general criterion of truth would be that which was valid of all cognitions without any distinction among their objects. But it is clear that since with such a criterion one abstracts from all content of cognition (relation to its object), yet truth concerns precisely this content, it would be completely impossible and absurd to ask for a mark of the truth of this content of cognition, and thus it is clear that a sufficient and yet at the same time general sign [Kennzeichen] of truth cannot possibly be provided. Since above we have called the content of cognition its matter, one must therefore say that no general sign of the truth of the matter of cognition can be demanded, because it is self-contradictory.[9]

A somewhat different formulation of this line of reasoning appears in Kant's logic lectures:

> A universal material criterion of truth is not possible; it is even self-contradictory. For as a *universal* criterion, valid for all objects in general, it would have to abstract fully from all difference among objects, and yet at the same time, as a material criterion, it would have to deal with just this difference, in order to be able to determine whether a cognition agrees with just that object to which it is related and not just with any object in general, in which case nothing would really be said.... [I]t is absurd to demand a universal material criterion of truth, which should abstract and at the same time not abstract from all difference among objects.[10]

The point is that a general and substantive criterion of truth would give rise to an irresoluble conflict between generality and particularity. Using 'theory of truth' for 'theory that provides a criterion of truth', we may express the radical disunity argument as follows:

(a) Truth consists in the particular agreement of a thought with its unique object.
(b) A general theory of truth must, in order to be general, abstract from the particularity of this relation.
(c) But a substantive theory of truth cannot (if it is to be substantive) abstract from its particularity. Hence:

---

[9] *Critique of Pure Reason*, 1st/2nd edition, 1781/87 (New York: Cambridge, 1998), A58–9/B83.

[10] *Lectures on Logic*, "The Jäsche Logic," 1800 (New York: Cambridge, 1992), p. 558.

(d) A general and substantive theory of truth is impossible.[11]

Now, on some level, this argument is persuasive. Not only is it formally valid; but its conclusion is, in some sense, correct: it is absurd to think that the theorist of truth could come up with a general and substantive criterion (or necessary and sufficient condition) that would determine, all by itself, the truth value of each and every truth bearer. But in a deeper and more important way the radical disunity argument is unsound. The argument assumes an altogether unreasonable conception of a substantive theory of truth as consisting of, or offering, a Kantian criterion of truth. There is no reason that a theory aiming at a philosophical explanation of truth be interested in, or be required to provide, such a criterion. Achieving a genuine understanding of truth does not mean detecting all the minute differences between any distinct truths, or determining what exactly has to be done in order to find out whether such sentences as 'Penguins waddle' and 'Snow is white' are true. There is a whole array of intermediate projects between the minimalist (deflationist) project of name clarification of truth and the maximalist project of providing a Kantian criterion of truth. And a substantive theory of truth aims at (some of) the intermediate projects. A comparison with the theory of knowledge might help. No one, least of all Kant, would require a substantive theory of knowledge to provide a full and detailed criterion of knowledge—a criterion determining, all by itself, with respect to each and every judgment, whether it should be included in our corpus of knowledge. Why should the theory of truth be required to provide such a detailed criterion of truth? One might answer that truth is special: the truth of a thought *is* dependent on its specific content and object. But does not the same hold for knowledge? (Are not the knowledge conditions of 'Penguins waddle' and 'Snow is white' just as particular and just as diverse as their truth conditions?) Should we conclude, then, that a substantive theory of knowledge is also impossible?[12]

---

[11] My reading of Kant's argument is different from James Van Cleve's in his *Problems from Kant* (New York: Oxford, 1999), chapter 13. Van Cleve regards Kant's argument as a version of the *diallelon*, an ancient argument which, based on circularity considerations, says that one cannot determine whether a cognition agrees with its object; one can only determine whether it agrees with another cognition of that object. While Kant's argument does appear in a section whose opening sentence mentions circularity—"The old and famous question with which the logicians were to be driven into a corner and brought to such a pass that they must either fall into a miserable circle [in the second edition, *Dialexe*; in the first, *Dialele*, that is, reasoning in a circle (editors note)] or else confess their ignorance, hence the vanity of their entire art, is this: **What is truth?**" [A57–8/B82]—attention to the *content* of his argument shows that it makes an altogether different point from the *diallelon*. I should indicate that I agree with other aspects of Van Cleve's analysis, for example, his claim (contra Putnam and Norman Kemp Smith) that this argument does not signal Kant's withdrawal from the correspondence view of truth.

The radical disunity argument is, in my view, best interpreted as a *reductio ad absurdum* of the criterial, maximalist view of a substantive theory of truth. Substantive theories are, in general, selective; they abstract from, that is, overlook, some aspects, features, and differences of objects in their domain.[13] The choice is not between a deflationist theory and a criterion of truth; the choice is between the former and a substantive account of the *major principles* and *facets* of truth.

It is not clear whether Kant himself drew a false dilemma between (what we call today) a deflationist theory and a criterion of truth. On the one hand, in introducing his argument he does appear to contrast the mere "name clarification" of truth (his conception of a deflationist theory) with a criterion of truth.[14] On the other hand, elsewhere he repeatedly affirms the existence of substantive accounts of various facets of truth and makes substantive claims about truth. For example, he declares that general logic provides a universal negative criterion of truth; he characterizes Transcendental Analytic as "a logic of truth" (saying it sets negative but not narrowly logical conditions on the possibility of truth); he claims that "transcendental truth, which precedes all empirical truth and makes it possible, consists in the general relation [of cognitions to the entirety of all possible experience]"; he suggests that an account of "the formal conditions of empirical truth" is possible; he proclaims that the principle of causality is a condition of empirical truth; he argues that the possibility of experience is a necessary condition for truth; he implies that a "sufficient mark of empirical truth" is possible; and he contends that his own theory, unlike Berkeley's, is capable of providing a "certain criterion for distinguishing truth from illusion" (in a more reasonable sense of 'criterion' than in the introduction to "Transcendental Logic").[15] Be that as it may, the dilemma raised by the radical disunity argument is a false dilemma. The radical disunity argument rejects the possibility of a substantive theory of truth by arguing against an absurd conception of such a theory.

---

[12] Note that for a while analytic philosophers did expect the theory of knowledge to provide something akin to a full and detailed definition, namely a necessary and sufficient condition of 'x knows that P'. But this is no longer the goal of (most) current epistemologists, nor was it Kant's goal. Note, too, that the chaotic results of the search for a necessary and sufficient condition of knowledge (reflected in the incessant stream of counter- and counter-counterexamples) undermined many philosophers' confidence in the possibility of a substantive analytic epistemology.

[13] One exception might be mathematics, whose objects are already abstract and incomplete (in Alexius Meinong's sense)

[14] See *Critique of Pure Reason*, A58/B82.

[15] See *Critique of Pure Reason*, A59–60/B83–84 and *Lectures on Logic*, 280–01, 455–56, 558–60, *Critique*, A62/B87 and A131/B170; A146/B185; A191/B236; A202/B247; A489/B517; A651/B679, and *Prolegomena to Any Future Metaphysics*, 1783 (New York: Macmillan, 1950), p. 124.

Furthermore, the radical disunity argument assumes that the theory of truth has only one goal: account for the truth conditions of sentences (cognitions, and so forth). But the theory of truth has other goals as well: explain the normativity of truth, determine its applicability to various fields of discourse, adjudicate between different conceptions of truth (correspondence, coherence, and so forth), elucidate the relation between truth and other topics of philosophical investigation, and so on. The radical disunity argument does not question the feasibility of any of these goals.[16]

But while the radical disunity challenge is a false challenge, the tensions between generality and particularity, unity and diversity, abstraction and detailed investigation, to which it directs our attention, do pose a genuine, if not insurmountable, challenge to knowledge. This challenge was taken up by Kant (in another section of the *Critique of Pure Reason*), James, and Wittgenstein, but I prefer to approach it through a more contemporary venue—the ongoing debate on disunity in science. It seems to me that many of the issues raised in this debate pertain to substantive theories in general, and the challenge facing the theorist of truth would be better understood by a judicious comparison with the one facing the scientist.

I.B. Moderate Disunity Challenge. Challenges to substantive theories of truth may be of three kinds: (a) challenges to substantive theories *in general*, (b) challenges to substantive *philosophical* theories, and (c) challenges to substantive theories of *truth*. The moderate disunity challenge falls under the first category, and to discuss it I will turn to the literature on disunity in science.[17]

The disunity of science is commonly conceived either as a disunity of theories or as a disunity of their subject matter, namely, nature.

---

[16] For further criticisms of the radical disunity argument, see Wright, "Truth: A Traditional Debate Reviewed," and Sher

[17] See, for example, P.W. Anderson, "More Is Different," *Science*, CLXXVII (1972): 393–96; Jerry Fodor, "Special Sciences," *Synthese*, XXVIII (1974): 77–115; Patrick Suppes, "The Plurality of Science," *Philosophy of Science*, II (1978): 3–16; John Dupré, *The Disorder of Things* (Cambridge: Harvard, 1993); Ian Hacking, "The Disunities of the Sciences," in Peter Galison and David Stump, eds., *The Disunity of Science* (Stanford: University Press, 1996), pp. 37–74; Nancy Cartwright, *The Dappled World* (New York: Cambridge, 1999); R.B. Laughlin and David Pines, "The Theory of Everything," *Proceedings of the National Academy of Sciences of the USA*, XCVII (2000): 28–31, and Laughlin, *et al*, "The Middle Way," pp 32–37.

These two types of disunity are interconnected, but the former places a greater emphasis on conceptual differences, the latter on "the disorder of things" (to borrow the title of John Dupré's book). The disunity of theories challenges the unification of science on three levels: (i) total unification (that is, construction of a "theory of everything"), (ii) intertheoretic unification (for example, reduction of psychology to biology), and (iii) intratheoretic unification (for example, elimination of the particle-wave duality in physics). The disunity of nature challenges the ability of science to systematize its subject matter. It is an open question how much order there is in nature and, as a result, whether nature can be subsumed under general laws. Now, it seems to me that both challenges can be generalized to other fields of knowledge, including philosophy with its broad, diverse, and highly complex subject matters—knowledge, ontology, meaning, and truth. Among the more general considerations raised by scientists and philosophers of science are:[18]

(1) *The complexity of the world:*
  (a) The world exhibits different complexities and interdependencies on different levels, and at each level of complexity entirely new properties appear (P.W. Anderson).
  (b) There are both higher and lower organizing principles, and in the course of investigation we sometimes have to add new levels of basic entities, concepts, and principles (Dupré, R.B. Laughlin and David Pines, Laughlin *et al.*).
  (c) The behavior of objects and properties is sensitive to a multiplicity of factors governed by multiple principles (Dupré, Nancy Cartwright).

(2) *The limitations of our cognitive powers:*
The great complexity of the world on the one hand, and our cognitive limitations on the other, limit our ability to comprehend it by a single, unified principle or theory (Comte[19]).

(3) *The partiality of knowledge:*
Human knowledge is, by its nature, partial: (i) universal principles and explanations are, due to their high level of abstraction and idealization, inherently partial, covering certain aspects of the phenomena under discussion while leaving others uncovered; (ii) as concepts and knowledge are extended, universal principles become partial, so that what was thought, at one point, to be complete,

---

[18] The points listed below constitute a pastiche of multiple ideas by multiple authors. At the end of each point I will indicate some of its sources. For fuller references, see the last footnote and the following footnotes.

[19] See Hacking, pp. 38–39.

turns out, later on, to be unfinished (Patrick Suppes, Noam Chomsky,[20] Ian Hacking, Stephen Hawking,[21] Laughlin and Pines).

(4) *Richness of interests, multiplicity of perspectives, human creativity.* There are many legitimate ways of dividing things into units and many advantageous points of view on things; these, given our creativity, give rise to a multiplicity of theories, a multiplicity which enriches our understanding rather than impedes it (Jerry Fodor, Suppes, Dupré).

(5) *Other methodological considerations:*
   (a) In any field of knowledge we are continuously confronted with new situations and new problems; as new problems arise, new theories and new methods of investigations are often required (Suppes, Hacking).[22]
   (b) There can be many unifiers, of different interest, in any field of knowledge; hence, there may be room for more than one unifying theory (Hacking).
   (c) Generality is not always a guide to a better theory. A small collection of simple principles is preferable to a single, highly complex, universal principle (Hacking).

None of these considerations has to do with the specific features of science, and all are either directly applicable or easily extendable to philosophy. In the field of truth, they have the potential of changing our perspective on existent theories, suggesting further developments, and pointing to new solutions to old problems. I will not be able to work out the details of these influences here, but a few examples might indicate the direction of change.

Consider the equivalence schema. Deflationists claim that the equivalence schema (or something like it) exhausts the topic of truth. But from the present perspective the equivalence schema describes only one, high-level, principle of truth, and as such it provides a *partial* account of truth, to be complemented by other accounts, centering on other principles of various levels of generality.

One example of a lower-level account of truth is given by Alfred Tarski's theory.[23] While the equivalence schema treats all sentences on a par, Tarski's theory distinguishes them along a given parameter—logical structure. Sentences exhibiting distinct logical structures have different Tarskian truth conditions, whereas those exhibiting the same logical structure have essentially the same truth condition. From the present perspective, Tarski's theory provides an account of a special factor of truth—the *logical factor*—and as such it is a mid- (or low-) level theory.[24] Most importantly, Tarski's theory, like the equivalence schema, offers a *partial* account of truth (logical structure is not the only thing that determines the truth value of sentences), and a "complete" account of truth must go beyond it.

---

[20] "Language and Interpretation: Philosophical Reflections and Empirical Inquiry" (1988), in John Earman, ed., *Inference, Explanation, and Other Frustrations* (Berkeley: California UP, 1992), pp. 99–128.

[21] See Hacking, p. 54.

[22] Partly a citation from Suppes, pp. 14–15.

[23] "The Concept of Truth in Formalized Languages" (1933), in *Logic, Semantics, Metamathematics* (Indianapolis: Hackett, 1983), pp. 152–278

One way of going beyond Tarski's theory was suggested by Field.[25] While Tarski's theory does not distinguish the truth conditions of atomic sentences in any informative manner, Field envisions an informative account of their truth conditions. The realization of Field's vision has, however, come upon great difficulties, and these are commonly attributed to a special feature of his approach, namely, its physicalistic orientation. (Field wants to base the account of atomic truth on a physicalistic theory of reference, application, and fulfillment—in short, satisfaction—for our nonlogical vocabulary.) The disunity perspective points to a different explanation. The problem, from our perspective, lies in an implicit assumption of Field's project, independent of physicalism, namely, that the satisfaction conditions of the entire vocabulary of human thought (minus a small part of this vocabulary—the logical vocabulary) are based on one and the same principle, or kind of principle. From our perspective this assumption is unwarranted: the totality of extralogical expressions exhibits an enormous diversity, and one cannot take it for granted that physical and moral expressions, biological and philosophical expressions, psychological and mathematical expressions, expressions pertaining to religion and expressions pertaining to technology, are all governed by the same satisfaction principle, or the same kind of principle.

Field's reaction to the difficulties faced by his project was a retreat to deflationism. But the right lesson from our perspective is not deflationism, minimalism, or quietism. A more productive lesson is openness and flexibility in devising our methodology. The theorist of truth, like the scientist, must adjust his methodology to the peculiarities of his subject matter, and to do so effectively he should take a cue from his fellow theorists and follow such straightforward, commonsensical, and workable guidelines as:

(A) In constructing a theory of truth, do not legislate in advance the form the theory shall take. Whether the study of truth will lead to the discovery of a single, universal principle (definition, schema, necessary and sufficient condition), or to a number of partial principles, is an open question, the answer to which largely depends on features of our subject matter, that is, on things that will emerge in the course of, not prior to, investigation. The answer also depends on our resources and capacities, but they, too, cannot always be determined at the outset of inquiry.

---

[24] For a relevant discussion of Tarski's theory, see Sher

[25] "Tarski's Theory of Truth."

(B) Do not think of the study of truth as focused on a single problem. Truth is a broad, complex, and diversified topic, and as such it poses a plethora of problems rather than a single problem. Today, it is common to center the study of truth on the subject of truth conditions, but although this undoubtably is a central subject, it does not exhaust, or even come close to exhausting the topic of truth. Other subjects include the normativity of truth, the role of truth in knowledge, the relation between truth and correspondence, skepticism and relativism with respect to truth, the interplay between mind and world in creating a standard of truth, and so on.

(C) In developing a theory of truth, aim at a fruitful balance between universality and particularity, similarity and diversity, abstraction and attention to detail, systematicity and applicability. Freeman Dyson's dictum that "every theory needs for its healthy growth a creative balance between unifiers and disunifiers"[26] applies not only to scientific but also to philosophical theories.

(D) Think of the development of a theory of truth as a dynamic process, in the course of which the theory is likely to expand, contract, undergo revision, change direction, stall, make leaps of progress, yield unexpected results, and so forth. The question is not whether the theory of truth *is* (atemporally or eternally) a unified or a disunified theory; the question is what steps can we take to increase its unity without sacrificing its substantiveness.

These guidelines suggest that we need not conceive the theory of truth as either a Tarski-style definition, or an equivalence-like schema, or a Kantian criterion. Tarski's characterization of truth with its emphasis on logical structure is extremely fruitful in logic, where it is incorporated in the definition of logical consequence and, through it, makes an invaluable contribution to logical semantics. But it is not clear that to understand whether, why, and how truth applies to, say, ethics we need a Tarskian definition of truth for moral discourse. A similar point applies to correspondence. Obviously, if we start with the usual paradigms of correspondence associated with simple observational statements ('Snow is white', 'Grass is green', 'The cat is on the mat'), we are likely to conclude that correspondence is out of the question in ethics. But if we investigate the moral domain without prejudice, if we take it by its own measure, we open up the possibility of new insights into correspondence, insights that would liberate us from the naive, simple-minded view of true thought as a *mirror of* (or as *isomorphic to*) reality.

---

[26] *Infinite in All Directions* (New York: Harper and Row, 1988), p. 47.

The disunity challenge, as a challenge to knowledge in general and/or to philosophy in particular, was discussed by a number of philosophers. To further clarify my view, I will indicate a few points of similarity and difference with the views of three of these: Kant, James, and Wittgenstein.

Kant. In "On the Regulative Use of the Ideas of Pure Reason" (first *Critique*), Kant says:

> To the logical principle of genera which postulates identity there is opposed another, namely that of **species**, which needs manifoldness and variety in things despite their agreement under the same genus, and prescribes to the understanding that it be no less attentive to variety than to agreement. This principle (of discrimination, or of the faculty of distinguishing) severely limits the rashness of the first principle (of wit [an innate talent of the mind for comparing and assimilating things that are superficially different]); and here reason shows two interests that conflict with each other: on the one side, an interest in the **domain** (universality) in regard to genera, on the other an interest in **content** (determinacy) in respect of the manifoldness of species.
>
> Reason thus prepares the field for the understanding: 1. by a principle of **sameness of kind** in the manifold under higher genera, 2. by a principle of the **variety** of what is same in kind under lower species.... We can call these the principles of the **homogeneity** [and] **specification**...of forms.
>
> The first law...guards against excess in the manifold variety of original genera, and recommends sameness of kind; the second, on the contrary, limits in turn this inclination to unanimity, and demands that one distinguish subspecies before one turns to the individuals with one's universal concepts.[17]

Like Kant, I regard the disunity problem as a problem of balance, but Kant's construal of the problem strikes me as too narrow. Kant tries to fit the disunity problem into a neat and orderly picture of our system of knowledge as a hierarchical, species-genera structure, but the tension between unity and diversity, as I see it, is more complex, intricate, and multidimensional than the species-genera picture suggests. Accordingly, I regard our system of knowledge as a polymorphic structure rather than a regular tree-structure, as envisaged by Kant.

---

[17] A654/B682 (explanation in square brackets based on editors' footnote 117, p. 749), A657–58/B685–86, A660/B688.

James. In his lectures on pragmatism, James subsumes the disunity problem under the classical puzzle of "the one and the many":

> If...we talk in general of our intellect and its needs, we quickly see that unity is only one of them.... What our intellect really aims at is neither variety nor unity taken singly, but *totality*. In this, acquaintance with reality's diversities is as important as understanding their connexion.
>
> The...point is to notice that the oneness and the manyness are absolutely co-ordinate here. Neither is primordial or more essential or excellent than the other.[28]

While James, too, regards the disunity problem as a problem of balance, his solution to the problem is fundamentally unbalanced. James recommends a shift toward radical pragmatism, a pragmatism that requires, at least in philosophy, rejecting the rational, the abstract, and the theoretical in favor of the experiential, the concrete, and the practical. I agree with James that balancing the demands of unity and diversity requires a certain amount of pragmatic "juggling." But I see no reason why this should conflict with rational, abstract, and theoretical reasoning. The disunity challenge is a challenge to the *design* of theories, not to theorizing itself; it necessitates the introduction of pragmatic considerations into philosophy, not the elimination of theoretical, rational, and abstract considerations from philosophy.[29]

Wittgenstein. Wittgenstein's disunity challenge is expressed in his "family resemblance" remarks. Wittgenstein objects to what I have elsewhere called "the myth of the common denominator," namely, the view that to understand a concept is to identify the common denominator of all objects falling under it, or to formulate a necessary and sufficient condition for falling under it (*op. cit.*). Many concepts, according to Wittgenstein, have no single defining characteristic (or a small set of such characteristics); instead, they exhibit "a complicated network of similarities overlapping and criss-crossing: sometimes overall similarities, sometimes similarities of detail."[30] This is the case, in particular, with concepts representing the traditional subject matters of philosophy: language, thought, number, and, significantly, truth.[31]

---

[28] "Pragmatism" (1907), in *Pragmatism and Other Writings* (New York: Penguin, 2000), pp. 1–132, citations from pp. 59 and 62

[29] A similar point was made by Robert Brandom with respect to Putnam's pragmatism—"Hilary Putnam: *Renewing Philosophy*," this JOURNAL, XCI, 3 (March 1994): 140–43. It should be noted that not all interpreters view James's pragmatism as constricting in the way I described (see, for example, Yemima Ben Menachem, "Pragmatism and Revisionism: James's Conception of Truth," *International Journal of Philosophical Studies*, III (1995): 270–89). But the tendency to associate pragmatism with a negative attitude towards abstract, rational theorizing is sufficiently prevalent to make it worthwhile to point out that an adequate solution to the disunity problem does not require such an attitude.

I share many of Wittgenstein's views: that philosophy tends to produce "network" rather than "one characteristic" concepts, that its theories face a serious disunity challenge, that its vulnerability to disunity is due to the breadth, complexity, and multidimensionality of the problems it seeks to resolve, and so forth. I also support Wittgenstein's conclusion that philosophers should increase the element of "looking" in their investigations: "Do not say: 'There *must* be something common, or they would not be called ['truths']'—but *look and see* whether there is anything common to all." "[D]o not think, but look!" (*op. cit.*, §66). Wittgenstein, however, goes too far. Like James he concludes that philosophy "may not advance any kind of theory," that it "must do away with all *explanation*," (*op. cit.*, §109)[32] and that it should abandon any aspirations to systematicity. These conclusions, I believe, are *nonsequitur*. It is an open question how much disunity there is in various branches of philosophy; whether this disunity rules out the existence of structure, hence of theory and explanation; and how resourceful philosophers will be in facing this disunity.[33] The tensions between unity and plurality, generality and particularity, abstraction and attention to detail challenge philosophy but need not stifle it. In fact, they create a fertile ground for theory construction, since what is it to construct a theory but to systematically connect elements that, prior to the construction, are disparate, varied, disorderly, and disconnected? Just as important, there is no real conflict between looking and thinking. Much of thinking is looking, and in many fields (for example, meta-logic) looking is abstract (for example, looking at a proof system to see whether it is complete). Philosophers ought to increase the element of "looking" in their methodology: carefully examine the objects of their inquiry, be open-minded and nondogmatic, aim at correctness, provide justification, be mindful of

[30] *Philosophical Investigations*, 3rd edition (New York: Macmillan, 1953), §66.

[31] See Juliet Floyd, "On Saying What You Really Want to Say: Wittgenstein, Gödel, and the Trisection of the Angle," in Jaakko Hintikka, ed., *Essays on the Development of the Foundations of Mathematics* (Boston: Kluwer, 1995), pp. 373–425

[32] See also §126. "Philosophy simply puts everything before us, and neither explains nor deduces anything."

[33] This point was also made by G.P. Baker and P.M S. Hacker, *Wittgenstein: Understanding and Meaning* (New York: Blackwell, 1980), p. 327.

counter-evidence, and so forth. But this does not conflict with theory and explanation.

The idea that truth might be based on multiple principles suggests a *moderate pluralism* with respect to truth. The idea is that the theory of truth may profitably be constructed as a *family* of theories, rather than a single theory. Each theory in the family would investigate some area, aspect, or factor of truth, and together these theories would produce (in the ideal limit) a comprehensive account of truth. This pluralism is moderate since, on the one hand, it does not rule out the possibility that a single, exhaustive, and substantive theory of truth can, in principle, be constructed; on the other hand, it holds that such a theory (though desirable) is not a *sine qua non* for a thorough and genuine (that is, substantive) understanding of truth.

Moderate pluralism has recently been advocated by a number of philosophers, for example, Wright and Lynch.[14] Wright, like me, argues that universal principles might not exhaust the topic of truth and if they do not, they may be complemented by other, more specific, principles. But for Wright this means that different fields of discourse (physics, mathematics, ethics, the comic, and so forth) might be governed by altogether different types of truth, say, physical discourse by correspondence truth and mathematical discourse by coherence truth. This is a rather radical division, and it is important to note that pluralism may come in more moderate versions. First, the plurality of truth may lie within the bounds of a single type of truth, say, correspondence. In that case, truth in all areas of discourse would be based on correspondence, but the principles underlying correspondence in physics would differ from those underlying correspondence in, say, mathematics. (Later on, I will argue that truth is in fact based on correspondence, and the potential multiplicity of principles of truth is indeed a multiplicity of correspondence principles). Second, the plurality of truth need not center on "fields" of truth; it might center on "factors" of truth. The point is that truth might be based on a multiplicity of factors, but the same factor may operate in diverse fields of truth. Thus consider the "moral" sentence 'All humans are good', and the "biological" sentence 'All humans are two-legged'.

---

[14] Other contemporary philosophers who consider pluralism with respect to truth (or cognate subject matters) are Davidson, Dummett, Resnik, Devitt, and Putnam, to name but a few. See Davidson, "Reality without Reference," *Dialectica*, XXXI (1977): 247–53; Dummett, "Realism," in *Truth and Other Enigmas* (Cambridge: Harvard, 1978), pp. 145–65; Michael Resnik, "Immanent Truth," *Mind*, XCIX (1990): 405–24; Devitt; and Putnam, "Sense, Nonsense, and the Senses: An Inquiry into the Powers of the Human Mind," this JOURNAL, XCI, 9 (September 1994): 445–517.

100
J 826

The truth values of these sentences are determined by (at least) three factors—the logical factor (which is reflected in the satisfaction conditions of the universal quantifier and the conditional), the biological or physical factor (which is reflected in the satisfaction conditions of 'is human' and 'is two-legged'), and the moral factor (which is reflected in the satisfaction conditions of 'is (morally) good'). Two of these factors are relevant to the truth value of both the moral and the biological sentence, and only one distinguishes between them. A pluralism based on factors of truth is finer and more nuanced than one based on fields of truth, yet it recognizes (and can account for) differences among truths in different areas of discourse.

Two objections to pluralist conceptions of truth naturally arise:

(1) If truth is based on multiple principles, in what sense are these principles principles of the same thing, namely, truth?
(2) If two statements are governed by different types of truth (or their truth is determined by different factors) how can logical inference transmit truth from one to the other? For example, how can a comical conclusion follow logically from a physical premise (or from a set of premises which essentially contains a physical premise)?[34]

Solutions to both problems will be offered in section II.

II. THE UNITY OF TRUTH

II.A. Unity Challenge. The unity challenge is the challenge of finding as significant, as comprehensive, as informative, and as enlightening unities as possible. Unity, as many philosophers have pointed out, is a condition as well as a goal of knowledge. One aspect of unity, namely, its contribution to the explanatory power of theories (a major mark of substantive theories), is especially relevant to the present inquiry. The connection between unity and explanation has been emphasized by several philosophers of science (Carl Hempel, Michael Friedman, Philip Kitcher, and others[35]), and much of what they say

---

[34] See, for example, Christine Tappolet, "Mixed Inferences: A Problem for Pluralism about Truth Predicates," *Analysis*, LVII (1997): 209–10, and "Truth Pluralism and Many-Valued Logics: A Reply to Beall," *Philosophical Quarterly*, L (2000): 382–85. A similar question was raised by Cora Diamond ("Unfolding Truth and Reading Wittgenstein," manuscript, 1999) with respect to the view that truth is relative to language games.

[35] Hempel, *Aspects of Scientific Explanation* (New York: Free Press, 1965), and *Philosophy of Natural Science* (Englewood Cliffs, NJ: Prentice Hall, 1966); Friedman, "Explanation and Scientific Understanding," this JOURNAL, LXXI, 1 (January 17, 1974): 5–19; and Kitcher, "Explanatory Unification," *Philosophy of Science*, XLVIII (1981): 507–31, and "Explanatory Unification and the Causal Structure of the World," in Kitcher and Wesley Salmon, eds., *Scientific Explanation* (Minneapolis: Minnesota UP, 1989), pp. 410–505.

KP2364

about the role, forms, and problems of unity applies to philosophy as well: philosophical unification, like scientific unification, may be local or global (unification of disparate elements of a single philosophical field, unification of hitherto distinct philosophical fields, unification of all fields of philosophy, and unification of philosophy with other fields of knowledge). It may center on elements of different kinds: laws and principles, theories, arguments, concepts, methods of inquiry, justification procedures, and so forth. Its goals, like those of scientific unification, may vary: decreasing the overall number of tenets and principles, creating reliable styles (patterns) of argumentation, harmoniously integrating disparate elements, and so forth. Philosophical unification may be stricter or looser: reduction versus supervenience, supervenience versus (mere) integration, hierarchical integration versus holistic integration, and so forth. Among the reductionist projects in philosophy are idealism, materialism, physicalism, logicism, the *Aufbau* project, the "linguistic turn," and extreme naturalism; all major philosophical systems and movements—rationalism, empiricism, transcendental idealism, moderate naturalism, and so on—aim at harmonious integration of some issues, principles, problems, and/or methods.

Two well-known pitfalls of scientific unification are *spurious* unification and *exclusionary* unification. Spurious unification trivially reduces multiple laws to fewer laws without gain in understanding. One example (due to Hempel and Paul Oppenheim[57]) is conjunction: unifying *A* and *B* by constructing their conjunction, *A&B*. In the field of truth, deflationists often define truth by an infinite list, or an infinite conjunction, of T-sentences (instances of the disquotational schema). Such a definition arguably provides a spurious unification of diverse truth conditions. Field's criticism of Tarski's theory is also one of spurious unification, directed at Tarski's list-like specification of the truth conditions of atomic sentences.

Exclusionary unification is a flaw in attitude: to think that the success of, say, string theory would leave no worthwhile scientific questions unanswered, or would rule out the usefulness of all other unifiers, is to yield to this pitfall. Disquotationalism arguably suffers from this flaw as well. It is not uncommon for a disquotationalist to say that *all* there is to truth is disquotation, meaning *all* worthwhile philosophical questions about truth are answerable by the disquotational schema, and the *only* genuine unifier of truth is this schema.

The connection between unity and explanation in science has been recently challenged by Margaret Morrison.[38] Morrison argues that the highest degree of unification is achieved by structural explanations, but the best scientific explanations are causal rather than structural. I will not get into the issue of structural versus causal explanation here. The claims that structural explanations are not the only unifying explanations, that different kinds of explanation are unifying in different ways, and that the most general explanations do not automatically impart the greatest degree of understanding, I find congenial. But the implicit suggestion that science admits only one type of effective explanation (namely, causal explanation), and the sweeping declaration that "general principles fail to be explanatory in any substantive sense" (*ibid.*, p. 33), I regard as unfounded. The latter, in fact, is refuted by Morrison herself, since many of her causal principles are in fact highly general. The search for a substantive theory, as I understand it, is first and foremost a search for explanatory unifiers, on various levels of generality.

---

[57] "Studies in the Logic of Explanation," *Philosophy of Science*, xv (1948). 98–115.

The view that in spite of its diversity truth can be unified by general principles is supported by Wright.[39] Wright proposes a series of universal principles of truth, including:

(P1) To assert is to present as true.
(P2) Any truth-apt content has a significant negation which is likewise truth apt.
(P3) To be true is to correspond to the facts.
(P4) A statement may be justified without being true, and vice versa.[40]

My own conception of the universal principles differs from Wright's in one respect. While Wright regards the universal principles as "platitudes," and as such as minimalistic, that is, deflationist, I see no reason why the universal principles could not be substantive (informative, explanatory, and so forth). The key to a balance between unity and diversity is, in my view, not platitudiness (with its implication of triviality), but *partiality*. The universal principles are partial in the sense of not exhausting the topic of truth, that is, leaving room for other principles, on various levels of generality.[41] Below I will propose two substantive and universal principles of truth: *Immanence* and *Logicality*. The former is related to Wright's (P1) and (P3), the latter is somewhat related to his (P2).

---

[38] *Unifying Scientific Theories* (New York: Cambridge, 2000).

[39] *Truth and Objectivity*, and "Truth. A Traditional Debate Reviewed."

[40] *Truth and Objectivity*, p. 34 (almost verbatim).

[41] It should be noted that Wright's claim that the unifying principles are minimal is made in a particular context, namely that of showing that truth aptness does not carry a commitment to a full-fledged realism. Construing the unifying principles as minimalistic is one way of showing this. It should also be noted that Wright does not rule out the possibility of substantive unifying principles. Thus he says, concerning the assertibility platitude: "On reflection,...[i]t is not necessary to insist that there is *no* suitable notion of deep assertoric content. It suffices that there is, at any rate, at least a more superficial one, carried by surface syntactic features; and that a minimal

II.B. Substantive Unifiers of Truth. 1. Two types of universal and substantive unifiers. Consider a broad and multidimensional concept, *C*, that (due to its breadth and multidimensionality) cannot be fully captured by either a single definition or a necessary-and-sufficient condition. This situation leaves open at least two possibilities for (substantive) universal unifiers of *C*: (i) a "core" unifier—a unifier that traces the roots of *C* to some general principle(s) of human thought and/or the world (without purporting to exhaust *C*), and (ii) a "specialized" unifier—a unifier centering on a particular factor of *C* (one among many) that, due to its special features, applies to all instances of *C* (or to many instances of *C* in every area to which *C* applies). Below I will delineate two universal and substantive unifiers of truth: a core unifier, *Immanence*, described by the "Immanence Thesis," and a specialized unifier, *Logicality*, described by the "Logicality Thesis."

2. Immanence Thesis. The Immanence Thesis (upper-case 'I') traces the roots of our concept of truth to three basic principles of human thought: *immanence* (lower-case 'i'), *transcendence*, and *normativity*.

A. Immanence. The immanence principle says that one basic mode of human thought is the *immanent* mode. By 'immanence' I understand something akin to, but more general than, W.V. Quine's notion of immanence.[42] Quine says that to speak immanently is to speak from within a theory, where speaking from within a theory is speaking literally and assertively, that is, declaring that a certain object or *n*-tuple of objects (mentioned in a sentence) possesses a certain property (also mentioned in it), or more generally that things are (literally) thus and so. For me, an immanent thought is not necessarily assertive or literal. A thought is immanent if it attributes some property to some object(s) or says (assertively or unassertively) that things are one way or another. In other words, any thought attachable to Frege's content stroke is immanent in my sense.

The Immanence Thesis says that truth is immanent in three ways: (i) truth is a property of immanent thoughts, (ii) the question of truth arises for all immanent thoughts, and (iii) truth statements—that is, statements of the form '*X* is true/false' (or 'It is true/false that *X*')—are immanent. Thus: (i) If truth is a property of *X*, *X* is a thought

---

truth predicate is definable on any surface-assertoric discourse"—*Truth and Objectivity*, p. 29.

[42] "Ontological Relativity," in *Ontological Relativity and Other Essays* (New York: Columbia, 1969), pp. 26–68.

that attributes some property to some object, or says how things are. ('Snow is white', for example, of which truth is a property, is a thought—a sentence—and it attributes some property (the property of being white) to some object (the stuff snow).) (ii) If *X* is an immanent thought, the question of truth, that is, the question 'Is *X* true?', arises with respect to *X*. (The question 'Is it true?' arises with respect to 'Snow is white', 'Coal is green', and so forth.) (iii) If *X* is a truth statement (for example, the statement '"Snow is white" is true'), then *X* attributes some property (namely, the property of being true) to some object (in our example, the sentence 'Snow is white').

The range of immanent thoughts is vast and diverse: immanent thoughts may take the form of a sentence or a theory, they may be realistic or fictional, contingent or law-like, directed at objects and properties of one kind or another (physical, mathematical, philosophical, moral), syntactically simple or syntactically complex (logical compounds, modal compounds, subjunctive conditionals), and so on. The vastness, diversity, and inner complexity of the immanent domain partly explain the breadth, complexity, and diversity of our concept of truth.

B. Transcendence. The principle of transcendence is a generalization of Tarski's principle of meta-linguistic predication. This principle says that a second basic mode of human thought is the transcendent mode: we transcend a given thought, or domain of thoughts, in order to *reflect* upon it. "Transcendence" has fallen into disrepute lately. To say that truth is transcendent is (so it is claimed) tantamount to saying that we have a "God's eye view" on the world, that we have access to "things in themselves," and so forth. *Transcendence, as I use this term here, has neither of these connotations.* What I mean by 'transcending' is casting a reflective look at a thought, or at a region of thought, from a standpoint external to it, yet within the domain of (human) thought. Transcendence does not rule out immanence. On the contrary, most transcendent thoughts are immanent: they attribute properties to thoughts, draw relations between thoughts, say this is how things are with respect to thoughts.[43] The Immanence Thesis says that truth is a transcendent property of thoughts and that truth statements are (immanent) transcendent statements.

C. Normativity. The principle of normativity says that a third basic mode of human thought is the normative mode, a mode closely related to the human proclivity for critical reflection. This is how Christine Korsgaard (who studies normativity in the moral domain, but draws analogies to other domains as well) explains it:

---

[43] For an earlier articulation of the notion of immanent transcendence, see Sher, "Is There a Place for Philosophy in Quine's Theory?" this JOURNAL, XCVI, 10 (October 1999): 491–524, p. 523.

> Normative concepts exist because human beings have normative problems. And we have normative problems because we are self-conscious rational animals, capable of reflection about what we ought to believe and do. That is why the normative question can be raised in the first place: because even when we are inclined to believe that something is right and to some extent feel ourselves moved to do it we can still *always* ask: But is this really true? And must I really do this?... It is...because we are normative animals who can question our experience, that normative concepts exists.... [I]t is always possible for us to call our beliefs and motives into question.[44]

In the present case, it is our disposition to question whether things are as our thoughts say they are, that leads to the concept of truth. The critical question is: 'Is it so as a given thought says it is?' and truth is a *standard* for a positive answer to this question. The question 'Is it so?' is an especially broad and basic question: it applies to any immanent thought, and is included in other critical questions (for example, critical questions concerning knowledge). The Immanence Thesis says that the concept of truth is a normative concept, and its breadth is associated with the universality of the critical question 'Is it so?' in the (vast) domain of immanent thought.

D. Immanence Thesis. Truth, according to the Immanence Thesis, lies at the juncture of three basic principles of human thought: immanence, transcendence, and normativity. Given an immanent thought, *t*, the critical, transcendent question 'Is it so as *t* says it is?' arises with respect to *t*, and truth is a standard for a positive answer to this question, a standard for *it being so as t says it is*: object *o* has the property *P*, objects $o_1,..., o_n$ stand in the relation *R*, and so forth. Truth, of course, is not the only standard for immanent thoughts; other standards (associated with other critical questions) include coherence, justification, empirical verification, explanatory value, utility, and so forth. But the critical question 'Is it so?' is one of the more basic questions of human thought, and truth, therefore, is a fundamental standard of thought. A theory of truth explains this standard, specifies its principles, and works out its connections to other standards of human thought.

In developing a theory of truth it is important to recognize that the domain of immanent thought is an ever expanding, ever changing domain. The dynamic nature of this domain is connected to our tendency to create new concepts, acquire new interests, develop new perspectives, raise new problems, revise our concepts and theories, devise new cognitive tools, and so on. One task of the theory of truth is to study how changes in the domain of immanent thought affect our concept of truth. Other tasks are to determine whether the concept of truth actually applies to a given domain of thought, how it applies to it (for example, directly or indirectly), and so forth. Expressivists, for example, argue that truth does not apply to ethics; Russell argues that truth applies to sentences containing definite descriptions indirectly, through sentences that do not contain such descriptions; physicalists argue that truth conditions for any statements must be reformulated in physicalistic terms; and so on. Much more remains to be said about the Immanence of truth and the three principles underlying it, but the main idea should be clear: truth is a transcendent standard, or a family of standards, for immanent thoughts—a standard for a positive answer to the critical question 'Is it so?' directed at such thoughts.

---

[44] *The Sources of Normativity* (New York: Cambridge, 1996), pp 46–49.

The Immanence Thesis is rich in consequences. Two of its consequences concern (a) correspondence, and (b) skepticism and relativism with respect to truth.

(a) *Correspondence.* Truth, according to the Immanence Thesis, is a standard for a positive answer to the question 'Is it so, as a given immanent thought says it is?' Given the content of this question, a positive answer carries us outside the given thought into things external to it, things it is about—"the world" in a broad sense of the word. The question is whether the objects the thought is about have the properties it attributes to them or, more generally, whether the world is as it says it is. This, of course, is a correspondence question, and in this way the Immanence Thesis implies a correspondence view of truth—the view that for an immanent thought to be true there must be some positive correlation between what it says (literally or nonliterally) and how things are. But while the Immanence Thesis affirms correspondence, the view it affirms does not suffer from the rigidity, simplism, and dogmatism that correspondence views are often charged with. The reason is that, as we have seen above, the thesis is sensitive to expansions, changes, and variations in the domain of thought, both internal and transcendent, and this sensitivity protects it from commitment to an overly simplistic, one-dimensional view of correspondence.

Correspondence, from the standpoint of the Immanence Thesis, is a *research program* rather than a *dogma.* Among the questions raised by this program are: What correspondence principles govern truth in logic? Mathematics? Physics? Psychology? (Or what are the logical, mathematical, physical, and psychological factors of correspondence?) What are their similarities and differences? Is there correspondence in ethics? What kind of correspondence? (Are we misled by the surface structure of ethical thoughts?) All these are open questions, and a judicious response to them requires an understanding not just of correspondence per se, but of the specific field of knowledge in which it is realized (to which its various factors belong). Below I will offer an account, or rather an outline of an account, of one type (aspect, factor) of correspondence, the *logical* one, based on prior investigations of logic.[45]

(b) *Skepticism with respect to truth.* Skepticism is often divided into two kinds: *local skepticism* and *global skepticism.* Local skepticism with respect to truth questions the existence of a standard of truth in a specific area; global skepticism questions the existence of a standard of truth in any area. Local skeptical challenges are part and parcel of a critical approach to knowledge; global skepticism is a barrier to knowledge. The Immanence Thesis affirms the legitimacy of skepticism, leaving the success of local skepticism (in particular cases) an open question, and posing counterchallenges to global skepticism.

Skepticism is sanctioned by the same principles that sanction truth itself: immanence, transcendence, and normativity. Given a "truth thought," *t*, that is, a thought that attributes truth/falsity to some thought (or truth aptness to some domain of thought), the critical transcendent question 'Is it so?' arises with respect to *t*, and one possible answer to this question is the skeptical answer: there is no way to determine whether things are as *t* says they are; there is no standard of truth for *t*.

Local skepticism. Consider the following local skeptical statements, expressing three degrees of skepticism with respect to *moral* truth:[46]

(1) Moral statements have a standard of truth, but this standard is a standard of psychological truth, not of *sui generis* moral truth. Explanation: although the surface structure of moral statements is '*X* has moral property *Y*', their deep structure is 'Speaker *Z* has

---

[45] See Sher, *The Bounds of Logic* (Cambridge: MIT, 1991); "Semantics and Logic," in Shalom Lappin, ed., *Handbook of Contemporary Semantic Theory* (Cambridge: Blackwell, 1996), pp. 509–35; "Did Tarski Commit 'Tarski's Fallacy'?" *Journal of Symbolic Logic*, LXI (1996): 653–86; "Is Logic a Theory of the Obvious?" *European Review of Philosophy*, IV (1999): 207–38; "The Formal-structural View of Logical Consequence," *Philosophical Review*, CX (2001): 242–61; and "Logical Consequence: An Epistemic Outlook," *Monist*, LXXXV (2002): 555–79.

[46] These statements are loosely based on discussions of moral expressivism in contemporary literature—see, for example, Allan Gibbard, *Wise Choices, Apt Feelings* (Cambridge: Harvard, 1990).

attitude *Y** toward *X*', where '*Y**' stands for a psychological attitude, mental state, or feeling correlated with *Y* (for example, '*Y*' stands for 'good' and '*Y**' for 'approbation'). The standard by which the truth of moral statements is measured is, therefore, a standard of psychological truth.

(2) Moral statements have a standard of truth (either a standard of *sui generis* moral truth or a standard of some other kind of truth), but this standard is not their primary standard of success. Explanation: while moral statements do attribute properties to objects (actions, intentions, and so forth), their main goal is to express, support, or arouse feelings or attitudes. Their primary success standard is therefore a standard of persuasiveness or expressivity, not a standard of truth.

(3) Moral statements have no standard of truth whatsoever, either primary or secondary. Explanation: moral statements are not immanent. They do not attribute any properties to any objects or purport to say how things are. They *merely* express attitudes or feelings, and expressions of feelings are not subject to a standard of truth.

The Immanence Thesis is compatible with all three degrees of moral skepticism, leaving their correctness an open question. That is, from the point of view of the Immanence Thesis it is an open question whether moral statements are literal, whether truth is their main standard of success, and whether they are immanent.

Global skepticism. Global skepticism says that there is no standard of truth for any thought whatsoever. The Immanence Thesis challenges the global skeptic to defend his view by showing either (1) that there are no immanent thoughts, or (2) that the critical transcendent question 'Is it so as *t* says it is?' does not apply to any immanent thought *t*, or (3) that there are no standards for a positive answer to this question with respect to any *t*. These challenges are not easy to meet. All of them involve negative existential claims which, especially in philosophy, are notoriously difficult to establish. Furthermore, each of them poses a special difficulty: to meet (1) the skeptic has to show that the skeptical statement itself is not immanent, and do so without using any other immanent statement—a self-defeating task according to many philosophers, for example, Thomas Nagel.[47] To meet (2) the skeptic has to explain how the question 'Is it so?' cannot apply to a thought that says that things *are* thus and so—something that goes against both grammar and semantics. And to meet (3) the skeptic has to establish not just that a certain truth standard does not apply to any thought, but that no truth standard at all (either one that has already been considered or one that might some day be considered) does—a formidable task indeed.

---

[47] *The Last Word* (New York: Oxford, 1997).

While the Immanence Thesis answers the first critical question posed at the end of section I—What makes the (potentially) diverse principles of truth principles of the same thing, namely truth?—in a traditional manner, that is, by pointing to a *core* principle underlying all truths, the Logicality Thesis offers a nontraditional answer to this question. The Logicality Thesis identifies a *highly specialized* factor of truth and shows that it is due to its specialty that it is universally applicable in the domain of truth. (Its specialty renders it so abstract that it overlooks all, or most, differences between fields of truth.)

3. Logicality Thesis. The Logicality Thesis says that one central factor of truth is *the logical factor*, a factor having to do with the role played by *logical structure* in rendering sentences true (false). The logical factor applies to truths and falsehoods in all areas of discourse and all fields of knowledge; at the same time, it is just one among a whole array of factors of truth, and its contribution to truth is specific and sharply delimited. The logical factor is, thus, an example of a factor of truth that, while substantive and universal, is partial and highly specialized—an example, in fact, of a factor whose substantiveness and universality are closely related to its partiality and specialization.

A comprehensive discussion of the logical factor in truth is beyond the scope of the present article.[48] Briefly, however, and without going into detail, my conception of the logical factor is the following:

The distinctive characteristic of the logical factor is its *formality*, which is most clearly expressed in the semantics of the logical constants. Logical constants are *formal* in the sense of *not distinguishing between objects (properties, relations) that are structurally the same*, or more precisely, not distinguishing between *argument structures* (see below) that are structurally the same. To see what this amounts to, consider three logical constants: identity ('='), the universal quantifier ('∀'), and conjunction ('&'). An argument structure for '=' is a pair, $<A, <b,c>>$, where A is a universe (that is, a nonempty set of individuals, either actual or possible) and b,c are individuals in A; an argument structure for '∀' is a pair, $<A,B>$, where A is as above, and B is a subset of A; and an argument structure for '&' is a pair $<A, <X,Y>>$, where either A is as above and X,Y are subsets of A, or A is a universe of propositions (with a truth value) and X,Y are members of A. (The former is the case when '&' functions as an operator on predicates

---

[48] For further discussion, see Sher, "On the Possibility of a Substantive Theory of Truth."

(open formulas); the latter is the case when '&' functions as an operator on propositions.) Let us call an argument structure whose universe is a set of individuals 'an objectual argument structure' and an argument structure whose universe is a set of propositions 'a propositional argument structure'. Say that (i) two objectual argument structures for a given constant are *structurally the same* if and only if they are *isomorphic*, that is, if and only if each is obtainable from the other by a 1-1 replacement of the members of their universes[49]; (ii) two propositional argument structures for a given constant are structurally the same if and only if they are "*T-isomorphic*," that is, their corresponding elements have the same truth values.[50] Now, '=' does not distinguish between isomorphic argument structures in the sense that if $<A_1,<b_1,c_1>>$ and $<A_2,<b_2,c_2>>$ are isomorphic, $<b_1,c_1>$ satisfies '=' in A_1 if and only if $<b_2,c_2>$ satisfies it in A_2. '∀' does not distinguish between isomorphic argument structures in the sense that if $<A_1,B_1>$ and $<A_2,B_2>$ are isomorphic, B_1 satisfies '∀' in A_1 if and only if B_2 satisfies it in A_2. And similarly for '&'. Since formality is preserved under combinations of logical constants, logical structures in general are formal.

The formality of the logical constants explains their universality, that is, their applicability in any field of discourse. If a logical constant, C, applies to an argument structure $<A,\alpha>$, where A is a universe of objects/propositions of any type (physical, psychological, cultural, mathematical, and so forth), then it applies to any isomorphic image, $<B,\beta>$, of $<A,\alpha>$, no matter what type the objects/propositions in B are. That is, if C applies to one type of object/proposition, it applies to all types of object/proposition. Moreover, the satisfaction conditions of C do not change from one type of object/proposition to another; that is, if $<A,\alpha>$ is isomorphic to $<B,\beta>$, the conditions under which α satisfies C in A are exactly the same as those under which β satisfies it in B. In short, C abstracts from all (nonformal) differences between domains of objects/propositions.

Now, generally the logical factor does not determine, all by itself, the truth value of sentences, but it combines with other factors to determine their truth value. For example, the truth value of 'Wet cats are funny',[51] formulated (in order to display its logical structure) as:

---

[49] Formally, the structures $<A,\alpha>$ and $<B,\beta>$ are isomorphic if and only if there is a bijection f from A onto B such that β is the image of α under f. For the origins of this criterion, see *The Bounds of Logic*, pp. 61–65.

[50] Formally, the propositional structures $<A,\alpha>$ and $<B,\beta>$ are T-isomorphic if and only if β is the image of α under some truth-preserving bijection f from A onto B (that is, a bijection f that assigns to every a in A a proposition b in B with the same truth value).

(1) $(\forall x)[(\text{Wet } x \,\&\, \text{Cat } x) \supset \text{Funny } x]$,

is determined by three factors—the physical factor, *P*, reflected in the satisfaction conditions of 'Wet' and 'Cat', the comical factor, *C*, reflected in the satisfaction conditions of 'Funny', and the logical factor, *L*, reflected in the satisfaction conditions of '∀', '&', and '⊃'.[52]

While the *L*-factor plays only a partial role in determining the truth value of (1), it determines, all by itself, the truth value of

(2) $(\forall x)(\text{Cat } x \vee \sim\text{Cat } x)$.

And in general, the *L*-factor is the only relevant factor in the truth (falsity) of logical truths (falsehoods). When we turn to logical inferences, the *L*-factor is the only relevant factor for their ***validity*** but not for their ***soundness***. Consider, for example, the logically valid and, let us assume, sound inference,

$$\begin{array}{ll} \text{(I)} & (1)\ (\forall x)[(\text{Wet } x \,\&\, \text{Cat } x) \supset \text{Funny } x] \\ & (3)\ (\exists x)(\text{Wet } x \,\&\, \text{Cat } x) \\ \cline{2-2} & (4)\ (\exists x)(\text{Cat } x \,\&\, \text{Funny } x). \end{array}$$

The validity of (I) is guaranteed by the *L*-factor alone, but the soundness of (I) is guaranteed by the *L*-factor ***together with*** the *P*- and *C*-factors.

We can now answer the second critical question raised at the end in section I: How can logical inference transmit truth across fields of discourse? For example: How can logical inference transmit truth from the physical domain to the domain of the comic?

The answer is: the logical factor plays a role in determining the truth value of sentences in all areas of discourse, including the physical and the comical, and this common role enables it to transmit truth from sentences in one area to sentences in another. More specifically:

(a) The truth value of sentences is generally determined by a multiplicity of factors, including factors that are not distinctive of the field of discourse to which these sentences belong.
(b) Due to its strong invariance property the logical factor can operate

---

[51] Tappolet, "Mixed Inferences: A Problem for Pluralism about Truth Predicates," p. 209.

[52] I use 'satisfaction' here as a generic term for 'application' and 'fulfilment', which apply to predicates, including quantifiers (2^{nd}-level predicates) and functions, respectively.

in all areas of discourse, that is, combine with all sorts of other factors in determining the truth conditions of sentences.

(c) The logical factor sets formal conditions on the truth of logically-structured sentences, and these conditions interact with other conditions, set by other factors, in determining the truth conditions of these sentences. For example, (3) is true if and only if the formal condition of nonemptiness is satisfied by the formal structure of intersection, applied to two collections of objects satisfying the physical conditions of being a cat and being wet, respectively.

(d) Due to connections between the formal elements of the truth conditions of sentences, the satisfaction of one constellation of formal conditions—that applicable to the premises of a given argument—may guarantee the satisfaction of another constellation of formal conditions—that applicable to the conclusion of this argument and, in so doing, tie the truth of the premises to that of the conclusion.

(e) In the case of (I), the logical factor guarantees that for any predicates, Φ,Ψ,Ω, and any factors, X, Y, Z, dominant in their satisfaction, the occurrence of the formal patterns

All objects which X-satisfy Φ and Y-satisfy Ψ also Z-satisfy Ω

and

Some object X-satisfies Φ and Y-satisfies Ψ

is always accompanied by occurrence of the formal pattern

Some object Y-satisfies Ψ and Z-satisfies Ω.

And it is this guarantee that enables us to infer the truth of the comic (yet logically structured) sentence (4) from the truth of the comic and physical (yet also logically structured) sentences (1) and (3).

Logical Correspondence. The Immanence Thesis implies that truth, in general, is based on correspondence. In what sense is logical truth (and logical consequence) based on correspondence? Our analysis suggests the following explanation: truth has to do with how things are in the world, and things in the world have formal properties in addition to physical, comical, and other kinds of properties. The two main principles of logical correspondence are: (i) Logical constants are denoting constants: they denote formal objects (properties, relations, functions) in the sense of formality explained above. (For example, '=' denotes the identity relation in any universe, '∃' denotes the second level property of nonemptiness (of 1st-level properties) in any universe, and '&', applied to predicates, denotes the operation of intersection, or more generally Cartesian product, in any universe.) (ii) Logical truth and consequence are based on formal laws governing and connecting structures of objects, the structures of objects (properties, relations) delineated by the sentences involved.[33] We can represent the workings of logical correspondence in the logical truth (2) and the logical inference (I) by the following diagrams:

$$\begin{tikzcd}[column sep=small]
 & \textbf{Language:} & \textit{Logically True} & `(\forall x)[Cx \vee {\sim}Cx]' \arrow[d] \\
\text{(2*)} & & & \\
 & \textbf{World (\textit{formal law}):} & \textit{Universal} \arrow[uu, Rightarrow] & C \cup \bar{C}
\end{tikzcd}$$

$$\begin{tikzcd}[column sep=small]
 & \textbf{Language:} & \textit{True}\ `(\forall x)[(Wx\&Cx)\supset Fx]', \arrow[d] & \textit{True}\ `(\exists x)[Wx\&Cx]' \arrow[d] \arrow[r, dashed, ""{name=U}] & \textit{True}\ `(\exists x)[Cx\&Fx]' \\
\text{(I*)} & \textbf{World (\textit{formal law}):} & W \cap C \subseteq F; & \textit{Non-empty}\ W \cap C \arrow[r, ""'{name=D}] & \textit{Non-empty}\ C \cap F \arrow[u]
\arrow[Rightarrow, from=D, to=U]
\end{tikzcd}$$

The truth of (2) is based on correspondence in the following way: (2) is true because (a) it attributes the property of being universal to the union of the set of cats and its complement (in the given universe of discourse), and (b) the union of these sets is in fact universal (in that universe). But (2) is not simply true, it is logically true. And its logical truth is due to the fact that the correspondence responsible for its truth is of a special kind, connecting the *logical structure* of (2) with a *law* governing the *formal* behavior of the properties denoted by its predicates.

The transmission of truth in (I)—assuming the premises to be true—is based on correspondence in the following way:

(A) Materially, the transmission of truth is due to the fact that: (a) the truth of (1) guarantees that the intersection of the set of cats and the set of wet things is (in fact) included in the set of funny things; (b) the truth of (3) guarantees that the intersection of the set of cats and the set of wet things is (in fact) not empty; (c) it is a regularity (that holds in the actual world) that whenever a nonempty intersection of a set of cats and a set of wet things is included in a set of funny things, the intersection of the set of cats and the set of funny things is not empty; (d) the nonemptiness of the intersection of the set of cats and the set of funny things (materially) guarantees the truth of (4).

(B) But the transmission of truth in (I) is not just materially, but also logically guaranteed. And its logical guarantee is due to the fact that the regularity mentioned in (c) is not accidental, but an instance of a *formal law*: the law that whenever a nonempty intersection of two sets is included in a third set, the intersection of the second and third sets is not empty.

---

[33] As I use the term 'law' here, laws contrast with accidental generalizations: a biological law holds in all biologically-possible structures of objects, a physical law holds in all physically-possible structures of objects,. ., and a formal law holds in all formally-possible structures of objects. Formal laws apply to all structures of objects, regardless of their nonformal properties

It is in this way that the truth of (2) and the validity of (I) are based on correspondence, not physical or comical correspondence, but logical, or formal, correspondence: a systematic connection between the logical structure of linguistic entities on the one hand and patterns of objects-possessing-properties-and-standing-in-relations that constitute formal laws on the other.[54]

III. CONCLUSION

In this article, I have examined the prospects of a substantive theory of truth and some of its challenges. Those who deny the feasibility of such a theory tend to construe its problems as relatively narrow, that is, as problems specific to the theory of truth (or at least to philosophy). My countersuggestion is that the problems facing the theory of truth are, in the first place, general methodological problems, problems arising for any theory of a broad, complex, and multifaceted subject matter in any field of knowledge, and that it is important to realize the universality of these problems in order to deal with them effectively.

I have concentrated on two interrelated problems facing the theorist of truth: unity and disunity. The challenge facing the theorist of truth, like that facing the scientist, is finding a fruitful balance between the unity and diversity of his subject matter—truth. That is, the theorist of truth must balance recognition of the multiplicity of principles of truth (the variety of ways truth is realized in different fields of knowledge) with a search for order, common principles, and systematic interconnections. The key to a balanced theory is, I suggested, partiality. We may think of the theory of truth as a family of theories, each investigating one central aspect or factor of truth, and all being connected by a network of unifying principles, on various levels of generality. The theory of truth, on this conception, is not a deflationist theory. Like other theories in other branches of knowledge it aims at a genuine and deep understanding of its subject matter, and this it tries to achieve by renouncing unhelpful preconceptions (for example, that of capturing the whole topic of truth by a single definition or schema) and by committing itself to an open-minded and undogmatic investigation. My approach to a substantive theory of truth involves a new, moderate version of pluralism. This pluralism subsumes all truths under the correspondence principle, but correspondence itself it construes as a network of interconnected (sub-)principles.

---

[54] For further discussion, see the works mentioned in note 45.

Turning to unity, I have emphasized the importance of unity for substantiveness (through explanation) and I distinguished two types of substantive unifiers: "core unifiers" and "specialized unifiers." The first type is exemplified by the Immanence Thesis, the second by the Logicality Thesis. The Immanence Thesis identifies a common source of our concept of truth in a combination of three basic principles of human thought: immanence, transcendence, and normativity. The Logicality Thesis identifies a special unifying factor of truth—the logical factor—a factor that, due to its unique features, partakes in the determination of truth in all areas of discourse regardless of their differences. Together, the Immanence and Logicality theses provide an answer to the two critical questions: (i) What is common to all truths? (ii) How is logical inference across diverse fields of truth possible?

Much work remains, of course, to be done. But I hope the general lessons from science, the idea of a moderate pluralism of correspondence principles, and the steps toward unraveling the substantive unifiers of truth, demonstrate (or make some progress toward demonstrating) the feasibility of a substantive philosophical theory of truth.

GILA SHER

University of California/San Diego

COMMENTS AND CRITICISM

HARDIN, TYE, AND COLOR PHYSICALISM

Larry Hardin has been the most steadfast and influential critic of physicalist theories of color over the last twenty years. In their modern form, these theories originated with the work of J.J.C. Smart and D.M. Armstrong in the 1960s and 1970s[1] and Hardin appropriately concentrated on their views in his initial critique of physicalism.[2] In his most recent contribution to this project[3] he attacks Michael Tye's recent attempts to defend and extend color physicalism.[4] Like Alex Byrne and David Hilbert,[5] Tye identifies color with the reflecting properties of objects ("reflectance physicalism"). Specifically, the determinate and determinable colors are identified with types of reflectances. (Setting some complications aside, the reflectance of an object is the proportion of light that it reflects at each wavelength in the visible spectrum.) These reflectance types are, in the terminology of Hilbert, *anthropocentric*—in the terminology of David Lewis,[6] they are not very "natural" properties—but whether objects have them does not depend in any interesting way on perceivers.

Although the focus is on Tye's position, Hardin's objections are clearly intended to apply to other versions of reflectance physicalism.[7]

---

[1] Smart, "On Some Criticisms of a Physicalist Theory of Colors," in Chung-ying Cheng, ed., *Philosophical Aspects of the Mind-Body Problem* (Honolulu: Hawaii, 1975), pp. 54–63; Armstrong, *A Materialist Theory of the Mind* (New York: Humanities, 1968), *Perception and the Physical World* (New York: Routledge, 1961).

[2] "Are 'Scientific' Objects Coloured?" *Mind*, xciii (1984): 491–500, "A New Look at Color," *American Philosophical Quarterly*, xxi (1984): 125–34, *Color for Philosophers: Unweaving the Rainbow* (Indianapolis: Hackett, 1988, 1st edition).

[3] "A Spectral Reflectance Doth Not a Color Make," this JOURNAL, c, 4 (April 2003): 191–202.

[4] *Consciousness, Color, and Content* (Cambridge: MIT, 2000). See also Peter Bradley and Tye, "Of Colors, Kestrels, Caterpillars, and Leaves," this JOURNAL, xcviii, 9 (September 2001): 469–87.

[5] Hilbert, *Color and Color Perception: A Study in Anthropocentric Realism* (Stanford: CSLI, 1987); Byrne and Hilbert, "Colors and Reflectances," in Byrne and Hilbert, eds., *Readings on Color, Volume 1: The Philosophy of Color* (Cambridge. MIT, 1997), pp. 263–88.

[6] "New Work for a Theory of Universals," *Australasian Journal of Philosophy*, lxxv (1983): 343–77.

[7] See Byrne and Hilbert, "Color Realism and Color Science," *Behavioral and Brain Sciences*, xxvi (2003): 3–21, and "Colors and Reflectances."

0022-362X/04/0101/37–43

However, as we shall argue below, Hardin's objections to reflectance physicalism in general are unsuccessful.

First, a remark on the dialectical situation. Hardin apparently takes for granted that reflectance physicalism has some independent motivation, and that the onus is largely on its critics to undermine it. However, he sometimes suggests that it is question begging to assume the truth of reflectance physicalism in replying to objections to it. For example, in discussing Tye's comparison of color illusions with shape illusions Hardin claims that there is a crucial difference between the two, namely that there is no way to measure color independently of perception while such methods do exist for spatial properties. The reflectance physicalist will reply that it *is* possible to measure reflectance—and so color—independently of perception. Hardin dismisses this reply, saying it assumes that "we already know that spectral reflectances are colors, and that is, once again, the point at issue" (195). But this runs together two kinds of arguments. If Tye's comparison with shape were being offered as part of an argument for the truth of reflectance physicalism then it would be question begging to assume at the outset that colors are reflectances. In the case in question, however, Hardin has claimed that reflectance physicalism is incompatible with certain facts about color appearances (see section I below). Obviously, assuming the truth of a theory is a bad way to argue for it; equally obviously, it is an essential part of showing the theory's compatibility with the empirical data. In what follows, we are merely trying to do the latter; we are not giving any positive argument for reflectance physicalism.

Second, a remark on the scope of Hardin's arguments. Although Hardin's official target is reflectance physicalism, his three objections are more inclusive. For example, the first objection (see section I below) is in fact directed against any theory of color that holds that an object's color typically does not depend on the colors of surrounding objects, and reflectance physicalism is just one such theory. The other objections are likewise directed at (different) broader targets. However, for simplicity we will largely ignore these complications.

Now to Hardin's objections.

I. CONTRAST

As Hardin nicely describes, the apparent color of an area of a scene depends not only on the physical characteristics of that area but also on the characteristics of the rest of the scene. One kind of dependency is illustrated by simultaneous contrast: the apparent color of a target area varies with changes in the surround.[8] Contrast effects are sometimes taken to motivate the view that an object has a color only relative to its surround.[9] Although a physicalist about color could endorse this, a *reflectance* physicalist could not, because the reflectance of an object is independent of its surround.

However, Tye's comparison of color contrast with some well-known illusions of spatial perception explains why contrast poses no threat to reflectance physicalism.[10] Illusions such as the Müller-Lyer illusion do not show that the length of an object is relative to its surround. Rather, they show that sometimes an object's surround can prevent us from seeing the length it actually has. Similarly, although often square objects look square, sometimes they do not—for example, when a square is superimposed on one side of a pattern of radiating spokes it looks wider at one end than at the other.[11]

Here it is helpful to draw a distinction between *perceptible properties*, on the one hand, and *conditions necessary for their perception*, on the other.[12] The two should not be confused, any more than the pressure of one's bicycle tires should be confused with the conditions necessary for detecting the pressure (the presence of a well-functioning pressure gauge, among other things). The conditions necessary for the (veridical) perception of shape include facts about the surround, and also facts about the perceiver. Plainly this does not show that shape itself—the property perceived—is in any way relative to the surround, or relative to the perceiver, or that no external object has a shape. Similarly, the conditions necessary for the perception of color include facts about the surround, and also facts about "mechanisms within the brain of the viewer."[13] But this does not show that color is relative to the surround, or to the perceiver, or that no external object has a color.[14]

Likewise, the fact that the apparent color of an object depends on its surround does not show that we always misperceive the object's real color. Tye, however, suggests that any influence of the background on perceived color generates a color illusion. According to Tye, "the colors things are *experienced* as having as a result of the contrast between the real color of the stimulus and the real color of the background are merely apparent.... Our experiences represent them as being instantiated when in reality they are not...contrast effects involve color illusions."[15] Tye here seems to be assuming that for veridical perception of color the apparent color of an object must not depend in any way on the reflecting characteristics of other objects in the scene. The problems with this view are many, as Hardin forcefully points out.[16] Fortunately for reflectance physicalism it is based on a misconception: for some colors, a certain kind of surround may simply be a condition necessary for their (veridical) perception.

---

[8] It is worth observing that many changes in background will produce little or no change in the perceived color of a target. Contrast effects are largest for very simple scenes involving backgrounds that are either highly saturated, very light, or very dark.

[9] See Jonathan Cohen, "Color Properties and Color Ascriptions: A Relationalist Manifesto," available from http://aardvark.ucsd.edu/~joncohen/color/relational.html; B.P. McLaughlin, "The Place of Color in Nature," in *Colour Perception: From Light to Object*, Rainer Mausfeld and Dieter Heyer, eds. (New York: Oxford, 2003), pp. 475–505.

[10] *Consciousness, Color, and Content*, pp. 153–55.

[11] As illustrated by Tye's figure 7 3 (p. 154).

[12] See Byrne and Hilbert, "Color Realism and Color Science," p 6.

[13] Hardin, "A Spectral Reflectance Doth Not a Color Make," p. 197

[14] See Hilbert, pp. 66–68.

But why is the surround important to the perception of color? Hardin himself has lucidly outlined an explanation of this in the opening part of his article. The light reaching the eye from an object—the color signal—is the product of the reflectance of the object and the spectral power-distribution of the illuminant. Even if we ignore the loss of information due to the small number of types of photoreceptors in the eye, the problem of estimating the reflectance of the object is still seriously underconstrained. Focusing on the achromatic colors for simplicity, a given color signal might be produced by a white object in dim illumination, a gray object in medium illumination, or a black object in bright illumination. Faced with this problem, the visual system employs various built-in heuristics: adaptation, and the use of information and assumptions about other objects in the scene before the eyes. One such assumption might be that the object reflecting the greatest amount of light is white, and that gray and black objects reflect certain fixed proportions of this amount—the darker the object, the smaller the proportion. It is not hard to see why, given these sorts of heuristics, the visual system would fail to detect that an object was black unless other lighter objects were in the scene.

One more remark before we leave the phenomenon of simultaneous contrast. The following line of reasoning might seem superficially cogent: nothing is black, because the visual experience of black always involves the "illusory" mechanism of simultaneous contrast. But this is a mistake. *Visual mechanisms* (for example, the mechanism of simultaneous contrast) are neither illusory nor veridical. Rather, it is the *output* of visual mechanisms—visual experiences—that are illusory or veridical. The same mechanism may produce illusory output on one occasion, and veridical output on another.

---

[15] *Consciousness, Color, and Content*, p. 156.

[16] In addition, Tye's assumption is in some tension with other elements of his discussion, namely the emphasis on the relative illumination-independence of apparent color. For reasons given by Hardin and discussed by us below, color constancy requires that the visual system make use of information about scene features other than the reflectance of the target object.

II. UNIQUE AND BINARY HUES

Hardin's next objection turns on the distinction between unique and binary hues.[17] There is a shade of green ("unique green") that is neither yellowish nor bluish, and similarly for the three other *unique hues*—red, yellow, and blue. In contrast, every shade of purple is a bit reddish and a bit bluish, and every shade of orange is a bit reddish and a bit yellowish. Purple and orange are accordingly *binary* hues. Hardin claims that "the unique-binary structure of the hues is essential to them and is not properly captured by any reflectance-based theory of colors" (198).[18]

This objection is certainly suggestive, but what exactly is it? Of course, there is a sense in which the reflectance physicalist can easily divide reflectance types into those that are unique, and those that are binary. The reflectance types that are unique are the ones identical to red, yellow, green, and blue; those that are binary are the ones identical to purple, orange, olive, and turquoise. Hardin, obviously, thinks that the physicalist owes us more. Hardin might be claiming that, if reflectance physicalism is true, the unique/binary distinction should be a "physically natural classification" of reflectance types. (Indeed, we have interpreted him that way in previous articles.[19]) However, this interpretation does not seem to be correct, because Hardin explicitly notes that Tye's theory fails to provide such a classification, and yet Hardin does not take this to be problematic.

It is not clear, then, what Hardin's objection is. Further, Hardin does not discuss the most developed physicalist-friendly accounts of the binary/unique distinction.[20] Pending further elaboration on Hardin's part, we can at least demand a Scotch verdict.

III. VARIATION

Although there is much agreement on what is green, dark-green, and yellowish-green, there is not much agreement on what possesses the highly determinate shades, like unique green. According to reflectance physicalism (and many other theories of color), some objects are unique green, and yet, with a fixed choice of viewing conditions, such objects will appear unique green only to a minority of normal perceivers. If, as Hardin says, "No scientific sense can be attached to the claim that some of the observers are perceiving the color of the stimulus correctly and others not" (200), then reflectance physicalism is false.[21]

---

[17] In Hardin's article, this is advertised as "The last major objection" (198). As explained in section III below, this is somewhat misleading.

[18] For Hardin's original version of the objection, see *Color for Philosophers*.

[19] Byrne and Hilbert, "Color Realism and Color Science," "Colors and Reflectances"; Byrne, "Color and Similarity," *Philosophy and Phenomenological Research*, LXVI (2003): 641–65.

[20] Byrne; Byrne and Hilbert, "Color Realism and Color Science"; Bradley and Tye.

(Hardin does not explicitly separate this objection from the one just discussed, but the two are quite different. The unique green objection, but not the unique/binary objection, crucially turns on color misperception.)

Suppose that a certain chip looks unique green to Fred in such-and-such viewing conditions. Let us grant, for the sake of the argument, that the facts of variation undermine Fred's perception-based reason to believe that the chip is unique green. (Note that this is consistent with Fred's knowing that the chip is *green*—and Hardin does not claim otherwise.) Generalizing, no one knows, of any object *o*, that *o* is unique green. (Note that this is consistent with our knowing that some objects *are* unique green—and Hardin does not claim otherwise.) Let us further concede, again for the sake of the argument, that this deficiency in our knowledge is an irremediable part of the human condition. In particular, the theory of perceptual representation will never advance to the point where we will have some "independent method," as Hardin says (199), for determining exactly which properties our color experiences represent objects as having.

Now unless we make some controversial verificationist assumption, it is hard to see why this scenario makes "no scientific sense," or is otherwise objectionable. Imagine, as an analogy, a population of intelligent, reasonably accurate thermometers. To one of these thermometers, the ambient temperature just "seems" to be such-and-such degrees Fahrenheit, as a chip might seem to Fred to be unique green. Like all measuring instruments, the thermometers are calibrated slightly differently. They all agree that the temperature right now is pretty high, around 70°F or so. But some think the temperature is 69°F, while others think it is 70°F, and yet others think it's 71°F. Some of them conjecture that being 70°F is a physical property of some kind, perhaps related to mean molecular kinetic energy. But the thermometers have no theory of intentionality that would enable them to establish conclusively that they are representing physical properties of this sort. And, since they do not have other ways of measuring temperature, they have no "independent method" of determining whether the temperature right now is exactly 70°F, or even whether it is pretty high. Still, some of these thermometers are perceiving the temperature correctly and others are not. Further, this lack of an independent method need not stop them from forming justified beliefs about the temperature. Perhaps none of them can justifiably believe that the temperature is exactly 70°F, but presumably they might justifiably believe that it is on the high side, or approximately 70°F.

---

[21] Ned Block has a particularly sophisticated reply to this argument that we will not consider here—see "Sexism, Racism, Ageism and the Nature of Consciousness," in Richard Moran, Alan Sidelle, and Jennifer E. Whiting, eds., *The Philosophy of Sydney Shoemaker, Philosophical Topics*, XXVI (1–2) (Fayetteville: Arkansas UP, 1999), pp. 39–70 According to Block, individual differences in color perception of the sort Hardin mentions are not differences in the colors things look to have, but are rather nonrepresentational differences in phenomenal character. A reflectance physicalist who agrees with Block may therefore claim that *all* "of the observers are perceiving the color of the stimulus correctly."

As for the intelligent thermometers, so for us. Despite presently—and, perhaps, eternally—lacking an independent method to determine the veridicality of our color experiences, we may be reasonably accurate detectors of the colors of things.

ALEX BYRNE

Massachusetts Institute of Technology

DAVID R. HILBERT

University of Illinois at Chicago

BOOK REVIEWS

The Problem of Perception. A.D. SMITH. Cambridge: Harvard University Press, 2002. i + 324 p. Cloth $47.50.

In perceptual experience, we seem to be presented with objects. Direct Realism[1] is the view that these presented objects are normal physical objects, realistically construed. Direct Realism is vulnerable to claims that accounting for the experiences of subjects of illusions and hallucinations requires us to take such subjects to be immediately aware of entities like sense data. Such entities, even if they are physical (for example, because identical with brain states), are not the normal physical objects of which we seem to be perceptually aware. Once these entities are brought in as objects of awareness in any experiential context, it is impossible to purge them from accounts of experience—we must conclude that in experience we are aware of normal physical objects only indirectly, if we are aware of them at all. Anyone defending Direct Realism will need to provide phenomenologically adequate accounts of both illusory perception and wholesale hallucination that do not attribute awareness of nonnormal entities to the subjects of such experience. A.D. Smith believes this can be done, if the view that perceptual sensory experience is *perceptual* because it is *conceptual* is warded off.

Smith argues that perceptual intentionality is *sui generis*: it cannot be understood on the model of the intentionality of thought, especially if the intentionality of thought is wrongly assumed to be fundamentally descriptive. Most theories of perception agree that perceptual consciousness differs from *merely* sensory consciousness in having intentional character. Smith wants to explain in virtue of what perception is intentional rather than simply noting that it is, and he wants his explanation to be compatible with the following three theses. First, perception is like mere sensation in being a fundamentally animal endowment; its intentionality cannot, therefore, be explained by appeal to concepts or to any distinctively cognitive activity. Second, the psychological states that are perceptions do not result from mere sensations being augmented by something from outside sensory consciousness. In order for this second thesis to be compatible with the

---

[1] Here, and with other capitalizations, I am following Smith's usage.

0022-362X/04/0101/44–53

first, sensory consciousness itself must have aspects that need not be present in every state of sensory consciousness, but which suffice for intentionality—and so perception—when they are. Third, because perceptual judgment is fundamentally *de re*, perception must serve to acquaint us with objects. (If sensations (whether "mere" or not) served only causally to occasion judgments, the contents of those judgments could be wholly descriptive.)

Smith wants, then, to present an account of perceptual consciousness as a sensory, intrinsically intentional, not-necessarily-conceptual state in which we are acquainted with normal physical objects, even when we are subject to illusions and hallucinations. In arguing that this account solves the Problem of Perception, Smith demonstrates that there is no easy way to maintain a commitment to Direct Realism while discharging the obligations to provide both a phenomenologically rich account of perception and an adequate account of *de re* thought in its relation to perceptual experience. The price Smith chooses to pay for these commitments is appropriately high.

The first part of Smith's book engages the Argument from Illusion while developing his view of perceptual experience. Anti-conceptualism is defended; a usefully wide range of historical views (for example, those of Thomas Reid and Arthur Schopenhauer) are discussed, and insights drawn from the work of Edmund Husserl and Maurice Merleau-Ponty are used to engage critically views propounded by writers in the analytic tradition. The second, shorter part of the book engages the Argument from Hallucination, making use of the account of perceptual experience defended in part I. While the discussion of the phenomenological issues raised by different sensory modalities is one of the most rewarding parts of the book, part II is more important, because Smith there makes a crucial modification to his account of the intentionality of perception. Because the introduction of this modification is rather like a plot twist I am loathe to reveal, I will first discuss the account of perception as it is developed in part I.

Smith begins by examining, for different sense modalities, what *phenomenologically* distinguishes mere sensation from perception in that modality. (An example is the distinction between visual perception and the visual sensations—the two-dimensional "light-show"—that can be produced by pressing on closed eye-lids.) Each phenomenological distinction turns out to be a version of one of three phenomena, which Smith concludes are the fundamental sources of perceptual experience: apparent spatial three-dimensionality, apparent movement of an organ of a sense modality in relation to the presented environment, and a phenomenon that, borrowing from Fichte, Smith calls the *Anstoss*, which is an awareness, in the tactile modality, of a check on one's activity. Each source of perceptual sensory experience has a feature Smith claims is *nonsensuous*—that is, not a matter of the *quality* of sensations—yet nevertheless intrinsic to experience. When a state in which sensory qualities inhere also involves one of these features, that state is perceptual and not merely sensory.

In the case of the *Anstoss*, Smith convincingly explains how one can be aware of an impediment to one's activity in the absence of sensation. Smith tackles the two other sources of perceptual consciousness by exploiting the Kantian point that in perceptual consciousness, but not in the having of mere sensations, a subject can distinguish between *a change of experience* and an *experience of change* in the object of experience. There is a phenomenologically accessible difference between changing sensations and unchanging perceptual appearances; the changing sensations manifest the subject's changing relationship to the apparently unchanging objects of her perception. These constancies, Smith argues, are actually definitive of the two non-*Anstoss* sources of perceptual consciousness: there is no way to have *apparent spatial three-dimensionality* apart from, for example, the possibility of objects appearing nearer or farther without appearing to change size; and the only way to sense *motion of a sense organ in relation to a sense field* is if the self-motion does not entail apparent motion or change in the object of awareness.[2] Smith concludes that for states of sensory consciousness that are wholly sensuous—where the *Anstoss* is absent—the presence of phenomenological constancies is not only sufficient but also necessary for perception.

Smith is now able to run an indirect argument against the Argument from Illusion. The Argument suggests that in illusion, a subject is directly aware not of a normal physical object but of a sensation. And because the object of the subject's illusory awareness is the same object as would be the object of her veridical awareness in a minimally different normal case, what goes for illusions goes for all perceptual states—subjects are immediately aware of sensations.[3] Now if, in a given state of sensory consciousness, our sensations were the objects of our awareness, then we would not be aware of either the phenomenological constancies (where we are aware of an apparently unchanging object *while* having changing sensations) or the *Anstoss* (where we are nonsensuously aware of a force, and aware of *that* force even if we mistake its direction or intensity). But it is subjectively accessible to us whether we are aware of one or more constancies, or the *Anstoss*; when we are so aware, any sensations—more precisely, any sensory qualia present in our conscious awareness—cannot be the objects of our awareness. So the only alternative suggestion presented by the Argument from Illusion is ruled out; we therefore have no reason to abandon our Direct Realist view that in both illusion and veridical perception we are aware of normal physical objects.

---

[2] In the absence of some phenomenological constancy, any sensory qualia in consciousness are not there through one of the fundamental sources of perception; the subject would not be in a perceptual state of consciousness. The materials Smith finds adequate to defeat the Argument from Illusion are necessarily present in consciousness whenever consciousness is perceptual.

[3] Smith claims this generalization can be avoided only by insisting that all apparent cases of illusion are actually cases of hallucination.

In part II, Smith addresses the Argument from Hallucination. This is a more serious challenge to Direct Realism, because any account of a hallucinating subject's consciousness seems to require taking him to be aware of *some* object—and yet in virtue of being a *hallucinating* subject, there is no relevant normal physical object available to figure in such an account. But preserving Direct Realism requires that no nonnormal object (like a sense datum) be introduced. The only obvious option is to deny that a hallucinating subject is aware of any object at all. Smith takes this to be so phenomenologically implausible that he would rather give up Direct Realism than accept this Extreme Proposal. Smith engages what he takes to be the best version of the Extreme Proposal—the one he finds in the work of Gareth Evans[4]—and sets out to demonstrate that the proposal is phenomenologically inadequate *and* actually unnecessary.

Smith's argument that the Extreme Proposal is unnecessary is complicated, not least because he wants to accept many of the premises that figure in the argument for it. Smith here endorses two important claims: (1) the intrinsic nature of a subject's conscious state is not fixed solely by phenomenological factors; and (2) it is not legitimate to infer, from the fact that two states have the same total, proximate (for example, neural) cause, that they have the same cognitive status. While Smith denies that perception is distinctively cognitive in any way that excludes animals as perceivers, he accepts that correct characterizations of genuine (veridical and illusory, but not hallucinatory) states of perception cannot leave out "the worldly object" of which the subject is aware (205).

Smith is also committed to the claim that perceptual judgments and perceptual awareness itself are fundamentally *de re*. And despite his anti-conceptualism, Smith holds that a thinking subject who is perceptually aware of an object is thereby in a position to entertain a *de re* thought about it. He acknowledges that a subject who could not entertain a *de re* thought about a given object cannot be said to be aware of that object (211). If a hallucinator cannot have *de re* thoughts about hallucinated objects, he is not aware of any objects qua hallucinator; and the Extreme Position maintains that he cannot.

---

[4] See *The Varieties of Reference* (New York: Oxford, 1982).

The linchpin of Evans's argument is that thoughts have understandable expressions. If a sentence is to count as the expression of a thought, something must count as understanding it. Smith argues against the view, which he attributes to Evans, that the relevant understandability is secured only if a hearer distinct from the speaker could understand the sentence. He claims one can be aware of an object—and so be in a position to have a *de re* thought about it—even if no one other than oneself could understand one's expression of that thought. A hallucinator could understand his own expressions of *de re* thoughts about his hallucinated object—and so count as having those thoughts—even though no one else could. We need not take the extreme view that the hallucinator is not aware of any object at all.

Smith's argument for the phenomenological inadequacy of Evans's account is seriously hampered by his delaying his treatment of its most effective aspect—its reliance on an account of pretense—until the end of the book, when he is considering last-ditch objections to his own radical proposal. Smith's reader is therefore misled about the extent of the phenomenologically relevant resources available to Evans, and perhaps more anxious for Smith's way out than she would otherwise be—especially as Smith seems to take the phenomenological argument to be the most important. Smith thinks he has a more phenomenologically satisfying account of hallucination than that offered by the Extreme Proposal, which in chapter 8 he presents as completely unable to give an account of what it is for a hallucinating subject "to *seem to confront* something" (225, italics original). As in part I, what is phenomenologically satisfying points the way to materials for escaping an attack on Direct Realism.

The escape itself is presented in the first few pages of chapter 9; the remainder of the chapter is devoted to explaining that the resultant view is not as absurd as it sounds. The chapter begins with a reminder that hallucination is a form of perceptual consciousness which is "phenomenologically identical to perception, being not only sensuous in character, but also intentionally directed to ostensible objects in a physical environment" (231). As we saw in part I, it is subjectively accessible whether the objects of awareness are sensations. Putting these two reminders together has the important consequence that when one hallucinates, one is not aware of sensations as objects.

(If one were, one would notice; and one would then have a means of detecting when one was hallucinating. But it is definitive of hallucination that a hallucinating subject need not be aware he is hallucinating.) If the hallucinating subject is, qua hallucinator, aware of some objects, but those objects are not sensations, what are they? The answer will prompt a reconsideration of everything Smith has claimed so far. The surprising solution starts from the claim that the nonsensuous elements (the phenomenological constancies and the *Anstoss*) that, when present in a state of sensory consciousness, suffice to make that state perceptual rather than merely sensory, are present whenever a state of sensory consciousness is *phenomenologically* perceptual. They are present, that is, *whenever* it seems to a subject that he is presented with an object. The conclusion Smith draws from this is worth quoting in full:

> it is the *same* non-sensory functions that we have seen to be present in veridical and illusory perception—where they give us, for all we have seen so far, direct access to the physical environment—that constitute the phenomenologically perceptual nature of hallucination. In both cases, such functions give us, not mere sensations, but an *intentional object* that transcends the flux of sensation...perceptions and possible hallucinations are phenomenologically identical.... Since they are thus identical, they have *the same kind of objects* for consciousness. Hallucination, equally with veridical perception, presents us not with sensations, or sense impressions, or sense-data, but with *normal objects*: normal *physical* objects, to boot, in the sense that they are presented in three-dimensional physical space.[5] The only way to answer the Argument is to say that the only difference between a veridically perceived object and a hallucinated object is that the latter does not exist, or is unreal (233–34, italics original).

The Argument from Hallucination is defeated because no nonnormal object needs to be introduced to do phenomenological justice to the experience of the hallucinating subject. And it is defeated in a way that vindicates Direct Realism: taking a hallucinator to be aware of a nonexistent intentional object does not prevent our taking a nonhallucinator to be aware of an existent intentional object. Smith's argument for taking veridical perceivers (whether they are subject to illusions or not) to be aware of real physical objects rests on his acceptance of externalism, glossed as the view that "the same-total-cause-same-total-effect principle does not apply to the cognitive or epistemic status of conscious states" (235). For the externalist, epistemic features of conscious states—for example, their being states of knowledge or not—are not exclusively determined by causal realities like the states of one's visual cortex. Smith now suggests that "facts concerning the *object* of perceptual awareness," as well as facts about the epistemic status of conscious states, escape the causal principle—they are not "exclusively to be analyzed in terms of the *realities* involved in the causal process" (235, first italicization added). Because hallucinated intentional objects are not realities, "the causal principle does not apply to them"—that is, the presence of a nonexistent intentional object in a hallucinatory state does not entail the presence of a nonexistent intentional object in a qualitatively identical nonhallucinatory state (236). Smith explains that although hallucinated objects are "qualitatively identical to real physical objects, they differ in status [they are not real]. It is precisely this sort of 'cognitive' or 'epistemic' fact that escapes the causal principle" (236).

---

[5] Although Smith's formulations in part I are careful, a reader is nevertheless likely to have construed "physical" differently from what is now revealed to be the construal required by Smith's official view. That is, "physical" applies to objects *presented* in space, not (or not simply) to objects that are, in G.E. Moore's useful phrase, to be *met with* in space—see "Proof of an External World," in *Philosophical Papers* (London: Allen and Unwin, 1959), pp 127–50.

Even if this argument succeeds, something distinctive about Smith's commitment to externalism emerges in these claims. Consider his statement about the different statuses of intentional objects. It is surely true that whether an object is real or not is a fact with epistemic import; it nevertheless seems odd to call it an "epistemic fact." Smith's deeming it one is related, I think, to his tendency to take externalism *simply* as an effort to avoid applying the same-total-cause-same-total-effect principle to epistemic matters. Of course, externalism must insist that epistemic facts outrun neural (or "narrow") facts. Avoiding inappropriate applications of the "causal principle" is necessary to externalism. But it would not suffice for everything that goes by the "externalist" label. One fairly common feature of externalism, and certainly of the externalism that interested Evans, is a concern with what causal relations between subjects and objects *contribute* to epistemic facts—the way genuinely epistemic facts sometimes depend on how a subject is causally embedded in her environment. This concern does not figure in Smith's externalism at all. When he discusses causal relations and processes, it is always—with two exceptions, discussed below—to caution against attributing the wrong kind of power to causation in the epistemic realm.

This raises a question about Smith's claim to have fully acknowledged the importance of *de re* thought. Of course, we should not fault Smith's externalism for failing to do what no externalism could: provide a *proof* of the existence of those objects that are, in G.E.

Moore's phrase, "to be met with in space." Even Evans's version of externalism simply holds that certain cognitive states are only possible when the subject of those states is in the right sort of causal relations to objects of those states. That we take ourselves to be in such states with some frequency is not a refutation of idealism. It only shows that *if* Evans is right about the requirements on, for example, demonstrative reference, then unless some form of realism is true, we are frequently mistaken about the true nature of our mental states.

Smith rejects the view that "it is not possible to make genuine reference to what does not exist" (248). And in rejecting the Extreme Proposal, he rejects Evans's view that demonstrative reference fails to occur in hallucination (250). When Smith discussed the Extreme Proposal in chapter 8, he presented Evans's endorsement of it as following from the views that (a) statements that express thoughts are in principle understandable by persons besides their speaker; and (b) *de re* statements are so understandable only when the *res* is a "public" object. This ignores a crucial aspect of Evans's thought: that *de re* thoughts expressed with demonstratives are immune to "errors of misidentification," and thoughts expressed with perceptually based demonstratives are so immune because of the subject's causal-yet-epistemically-relevant link (what Evans calls an "information-link") to the object referred to. That is, it is essential to the perceptually-based *de re* thoughts in which Evans is interested that their objects be real.

It is not until late in Smith's project—his effort, in the last portions of chapter 9, to portray intentionalism as an ontologically respectable view—that it is fully clear that the notion of the *de re* that has been so important throughout the whole book is *wholly* phenomenological. What interests Smith is the subject's *sense* of her thought as *de re*. He is not interested, as Evans is, in the subset of those thoughts with what Smith calls "*de re* character" in which the subject is thinking of an object in virtue of her (epistemically significant) causal links to it. A "link" is a relation, and an information-link is a relation to a real object. But Smith's intentionalism denies *both* that perceptual awareness is a relation to a real object (the object could be nonexistent) *and* that perceptual awareness is a relation.

Smith reassures us that proper intentionalism is ontologically reductive: "[n]onexistent intentional objects supervene on intentional experiences" (244). Nevertheless, reference to intentional objects is *necessary* to descriptively adequate accounts of phenomenologically perceptual experience. If we press on the nature of this necessity, and on the question of what reference to a nonexistent intentional object amounts to, the contrast with the Extreme Position, at least with respect to phenomenological adequacy, looks less clear than

Smith takes it to be. Evans clearly holds that the *only* way to characterize the experience of a hallucinating subject is to say that it *seems* to him *as if* he is confronting a real object. Smith finds this phenomenologically inadequate primarily because he does not find in it any *explanation* of *why* it so seems to the subject. Smith's account offers the explanation that it so seems because he *is* confronting an object—though the object only *seems* to be real. But since Smith does not construe "aware of" relationally, "confront" must also be an experiential characterization with no implication of relationality. Smith can still offer the explanation that the experience has the phenomenally confrontational character it does because it involves the nonsensuous elements whose presence in sensory consciousness makes that consciousness phenomenally perceptual. But it is not clear that this account has dramatically more explanatory power than the account an adherent of the Extreme Proposal could give—namely, that the consciousness of the hallucinating subject is properly described as one in which he *seems* to be perceptually aware of an object.

Smith's concluding arguments address the worry that intentionalism causes problems for veridical perception (where the object of perceptual consciousness is real). Intentional objects supervene on intentional experiences; real objects do not. Smith maintains that phenomenology wholly determines intentional objects. The explanation of why the Direct Realist ought nevertheless accept the intentionalist account of perception is roughly as follows: when the intentional object is nonexistent, characterizations of the object cannot go beyond what is "sensorily given"—any claim that the object is *F* requires that *F* be a sensible quality. But when the intentional object exists there is "ontological grounding" for "extensional" perceptual attributions—that is, one may claim that the subject, for example, sees an object that is *F*, where *F* need not be a sensible quality (264).

This is the framework Smith uses to account for the fact, mentioned in the book's last paragraph, that we stand in real-world relations with "entities that *focally govern* the course of our sense-experience" (271, italics added). In veridical perception, the intentional objects delivered by perceptual consciousness are identical to "the normal physical entities with which we stand in real-world relations" (271).[6] Even if

---

[6] I found only one other case of positive reference to real (read as "causally significant") relations. When Smith first presents the worry that veridical perception is, for the realist, awareness of "real-world, physical objects," he notes that in accounts of veridical perception, "an object is allocated to a perceptual state at least partly on the basis of non-phenomenological factors: real-world factors involving the *real relations* between a perceiving subject and elements in his *physical* environment" (257, italics added) Here, presumably, "physical" does not simply mean: presented in phenomenal three-dimensional space.

Smith can account for these facts, it should be noted that intentionalism in the philosophy of perception does not take subjects' real-world relations to their real physical environment as a serious starting point for reflection. And some Direct Realists might complain that a defense of Direct Realism against arguments from illusion and hallucination should begin with those relations. This complaint is *not* the simple-minded complaint that externalism failed to prove Realism. The concern is rather that, regardless of whether Realism can be proved or not, the intentionalist defense alters any Realism it defends. Unless the defense alters Realism beyond recognition, this is not itself reason to reject intentionalism. But it does raise the question of whether the high price of intentionalism is worth paying for the type of Realism intentionalism can support. Smith's rhetorical strategy of deferring his discussion of intentionalism to the end of the book puts the reader in the position of wondering whether the trade offs she was convinced by Smith to make are fully compatible with that version of Direct Realism for whose sake she was ostensibly making them. If they are not, that would complicate, though not diminish, the significant achievements of *The Problem of Perception*, which is a fascinating, comprehensive, and tightly argued book.

MAURA TUMULTY

Johns Hopkins University

MEMO TO AUTHORS

The Journal of Philosophy is published monthly, in issues that (ordinarily) contain only 56 pages of copy; this limits the length of papers the editors can normally publish to no more than 7,500 words, or 18 *Journal* pages. A more precise way of determining length is by characters (including spaces): 2,850 characters per *Journal* page, or 51,300 characters per *Journal* article of 18 pages.

The editors request that three copies of all manuscripts submitted be prepared according to the following guidelines:

(1) All material, including the quotations and footnotes, must be typed, double-spaced, and one-sided, with ample margins left and right, top and bottom.
(2) No bibliographies will be published; bibliographical material should be put into footnotes.
(3) Footnotes should be complete, according to *Journal* style (see any issue). Journal citations should include: name of journal in full, volume number, issue number, month or season, year, and pages. Book references should include: publisher, U.S. city of publication, and date. Footnotes should be gathered at the end of the article.
(4) Quotations should be checked by the author; capitalization, italics, spelling, and the like should match the original.
(5) If there are diagrams, they should be provided camera-ready by the author. If there are special symbols, they should be widely spaced and clearly marked. Standard logic symbols are available, and letters of the Greek and script alphabet.
(6) No underlining; please limit use of italics to technical terms (when they are introduced) and to variables.
(7) No contractions.

The editors are willing to read and evaluate a manuscript that does not meet these guidelines, provided the author agrees to meet them before submitting a final manuscript. The editors may ask for revision of excessively long manuscripts, and may require reformatting of accepted manuscripts. They also request an electronic version with the final hard copy.

The Journal of Philosophy does not publish papers that are primarily historical or expository; the editors urge that such manuscripts be submitted elsewhere. While we do not publish manuscripts that mainly comment on articles published elsewhere, we do have a "Comments and Criticism" section in which short responses to *Journal* articles may be published.

Manuscripts may be sent to: John Smylie, Managing Editor, *The Journal of Philosophy*, 709 Philosophy Hall, 1150 Amsterdam Avenue, Mail Code 4972, New York, NY 10027.

The Illusions of Egalitarianism

John Kekes

"*The Illusions of Egalitarianism* is an excellent and much-needed criticism of the egalitarianism that dominates contemporary Anglo-American ethics and political philosophy John Kekes is devastating in showing how egalitarian thinkers attempt to impose an inappropriately theoretical template on the real world which highlights just one moral consideration above all others and is highly intolerant of diversity or world pluralism."—Anthony O'Hear, Honorary Director of the Royal Institute of Philosophy. $29 95

The Unraveling of Scientism

American Philosophy at the End of the Twentieth Century

Joseph Margolis

"Joseph Margolis takes us deep into the trenches and high above the battlefields of contemporary analytic philosophy to offer a big-picture perspective of scientism's last stand. He projects a future greening of philosophical analysis that preserves conceptual rigor but discards mainstream commitments to mind-body reductivism and semantic extensionalism."—Dale Jacquette, The Pennsylvania State University $35 00

Vico's Uncanny Humanism

Reading the "New Science" between Modern and Postmodern

Sandra Rudnick Luft

"Relying on some postmodern as well as rabbinical insights, Luft finds in Vico's work a vindication of poetic creativity outside the parameters of traditional humanist and theological interpretations focused on subjective agency Leaving behind the modern subject and its so-called 'end,' Vico emerges as the champion of a transformative praxis appropriate for our time."—Fred Dallmayr, Notre Dame. $45 00

Race, Racism, and Reparations

J. Angelo Corlett

"Until now, most philosophy texts on race have focused narrowly and monochromatically on the black experience This challenging new book broadens the traditional spectrum to include the neglected browns and reds, Latino and Native American colors, of the American racial palette. The result is a much richer picture of the moral complexities of the ethnic and racial landscape, from the subtleties of how best to analyze Latino identity to the highly contentious issue of reparations to Native Americans."—Charles Mills, author of *The Racial Contract*.
$47.50 cloth, $19.95 paper

The Transmission of Affect

Teresa Brennan

"In this rich and provocative book, Teresa Brennan has developed an alternative theory of affects that challenges any notion of self-contained individuals. Her theory of the transmission of affect insists on a constant communication between individuals and their environments, with social pressure acting at the intersection of the social and the material to bring together mind and matter through energetics"—Kelly Oliver, author of *Witnessing. Beyond Recognition* $45 00 cloth, $19 95 paper

Sovereign Nations, Carnal States

Kam Shapiro

"*Sovereign Nations, Carnal States* is a systematic study of the very important issue of 'somatics' in political action Kam Shapiro chooses his analytic cases—Augustine, Hegel, Nietzsche, Benjamin, Schmitt, and Derrida—very wisely, showing why practices of embodiment always complicate and qualify the workings of politics."—Timothy W Luke, Virginia Polytechnic Institute and State University.
$49 95 cloth, $19 95 paper

Cornell University Press • **www.cornellpress.cornell.edu**

New from The MIT Press

Seeing and Visualizing
It's Not What You Think

Zenon Pylyshyn

"I emphatically recommend the book not only to the specialist, but to anyone wanting an up-to-date text on the topic generally." — Georges Rey, University of Maryland

Life and Mind: Philosophical Issues in Biology and Psychology series
A Bradford Book
592 pp., 116 illus. $50

Consciousness and Persons
Unity and Identity

Michael Tye

"Tye presents a compelling, original, philosophically sophisticated theory of the unity of phenomenal consciousness without ever getting bogged down in philosophical trench warfare." — Brian P. McLaughlin, Rutgers University

Representation and Mind series
208 pp., 11 illus. $35

Radiant Cool
A Novel Theory of Consciousness

Dan Lloyd

"Dan Lloyd may not have solved the mystery of consciousness, but he has done something nearly as difficult: he has turned the quest for that solution into a page-turner. Ecstatically playful, *Radiant Cool* breaks new ground in the genre of philosophical fiction." — Rebecca Goldstein, author of *The Mind-Body Problem*

256 pp., 14 illus. $24.95

Natural Ethical Facts
Evolution, Connectionism, and Moral Cognition

William D. Casebeer

"Integrating evolutionary biology and connectionism with Neo-Aristotelian and Deweyan ethical theory, Bill Casebeer makes a powerful case for the scientific naturalization of ethics." — William Rottschaefer, Lewis and Clark College

A Bradford Book • 224 pp., 2 illus. $35

Preference, Belief, and Similarity
Selected Writings

Amos Tversky
edited by Eldar Shafir

"This collection offers the best of Tversky, the best of the best. It is amazing how many of these articles are already classics, not only in the fields of choice and decision making, but in psychology in general." — Edward E. Smith, University of Michigan

A Bradford Book
850 pp., 158 illus. $60 paper

Reasoning About Uncertainty

Joseph Y. Halpern

"An authoritative panoramic reference for philosophers, cognitive scientists and artificial intelligence researchers." — Judea Pearl, University of California, Los Angeles

456 pp., 12 illus. $50

now in paperback

Reasoning About Knowledge

Ronald Fagin, Joseph Y. Halpern, Yoram Moses, and Moshe Y. Vardi

"It is easy to foresee that this book will become a classic." — Fabrizio Sebastiani, *The Computer Journal*

491 pp., 13 illus. $39.95 paper

now in paperback

The Illusion of Conscious Will

Daniel M. Wegner

"A remarkable demonstration of how psychology can sometimes transform philosophy." — *The Times Literary Supplement*

A Bradford Book
419 pp., 57 illus. $17.95 paper

To order call **800-405-1619.**

http://mitpress.mit.edu

The Puppet and the Dwarf

The Perverse Core of Christianity

Slavoj Žižek

"With this book Žižek consolidates his reputation as the foremost intellectual gadfly of the postmodern cosmopolis." — Richard Wolin, City University of New York

Short Circuits series • 190 pp $16.95 paper

The Shortest Shadow

Nietzsche's Philosophy of the Two

Alenka Zupančič

"Against the postmodernist domestication of Nietzsche's philosophy, Zupančič restores the shock of Nietzsche's style and thought, reading him alongside Jacques Lacan and Alain Badiou as a revolutionary 'philosopher of the event.'"
— Christoph Cox, Hampshire College

Short Circuits series • 190 pp. $16.95 paper

Reflections and Replies

Essays on the Philosophy of Tyler Burge

edited by Martin Hahn
and Bjørn Ramberg

Essays by various philosophers on the work of Tyler Burge and Burge's extensive responses.

424 pp $45 paper

What Is Thought?

Eric B. Baum

"Presents a novel thesis—that the mind is a program whose components are semantically meaningful modules—and explores it with a rich array of evidence drawn from a variety of fields." — David Waltz, Columbia University

A Bradford Book • 495 pp., 34 illus. $37

Time, Tense, and Reference

edited by Aleksandar Jokić
and Quentin Smith

Original essays by philosophers of language and philosophers of time exploring the semantics and metaphysics of tense.

488 pp , 8 illus. $38 paper

Truth and Justification

Jürgen Habermas
translated by Barbara Fultner

Essays by Jürgen Habermas on truth, objectivity, normativity, naturalism, and realism after the linguistic turn.

368 pp $40

The Letters of George Santayana Book Six, 1937-1940

The Works of George Santayana, Volume V, Book Six

George Santayana
edited and with an Introduction by William G. Holzberger

The sixth book of the letters of American philosopher George Santayana, covering the years 1937 to 1940.

688 pp $60

now in paperback

Seeing Double

Shared Identities in Physics, Philosophy, and Literature

Peter Pesic

"Pesic suavely creates a masterpiece by saying much in little space."
— *Booklist*

192 pp , 11 illus. $14.95 paper

now in paperback

Interpreting Minds

Radu J. Bogdan

"Radu Bogdan has made a provocative contribution to the broad field in which philosophy, psychology, and biology overlap."
— *Contemporary Psychology*

A Bradford Book • 318 pp. $27.95 paper

now in paperback

Minding Minds

Radu J. Bogdan

"This is an interesting and stimulating book. The main idea that interpretation is part of the human mind is very appealing and well supported . . . Thought provoking and enriching."
— *APA Review of Books*

A Bradford Book
231 pp , 17 illus. $21.95 paper

Philosophy New from Chicago

Camus and Sartre

The Story of a Friendship and the Quarrel that Ended It

Ronald Aronson

"A major contribution to intellectual history. Aronson gives a thorough and detailed description of the background of both writers from a biographical and an ideological point of view, and he also provides the best analysis available of the 'quarrel' between Sartre and Camus."—Michel Rybalka, author of *The Writings of Jean-Paul Sartre*

Cloth $32.50

Nietzsche and Music

Georges Liébert

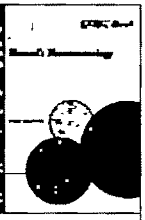
Translated by David Pellauer and Graham Parkes

"*Nietzsche and Music* . . . makes a real contribution to our understanding of this controversial genius, and shows us the extent to which his life and writings were a kind of commentary on another, more congenial but, alas, imaginary person, Nietzsche the musician."
—*Times Literary Supplement*, on the French edition

Cloth $38.00

The University of Chicago Press

ELEMENTS OF PHILOSOPHY



The Crossing of the Visible

JEAN-LUC MARION

Translated by JAMES K. A. SMITH

Ranging across artists from Raphael to Rothko, Caravaggio to Pollock, *The Crossing of the Visible* offers both a critique of contemporary accounts of the visual and a constructive alternative. According to Marion, the proper response to the 'nihilism' of postmodernity is not iconoclasm, but rather a radically iconic account of the visual and the arts which opens them to the invisible.

$16.95 paper $40.00 cloth

Elements of Ethics

ADRIAAN T. PEPERZAK

This book is an unorthodox attempt at renewing the basic questions and principles of philosophical ethics. It focuses on the descriptive and conceptual analysis of the experiences through which human lives become aware of themselves as being provoked and urged to respond appropriately to the various dimensions and phenomena of the universe.

$24.95 paper $60.00 cloth

Husserl's Phenomenology

DAN ZAHAVI

Drawing upon both Husserl's published works and posthumous material, *Husserl's Phenomenology* incorporates the results of the most recent Husserl research. It can consequently serve as a concise and updated introduction to his thinking.

$15.95 paper $45.00 cloth

Stanford

800.621.2736 | [illegible]

Philosophy New paperbacks from Chicago

The Work of Mourning

Jacques Derrida

Translated and Edited by Pascale-Anne Brault and Michael Naas

"Among those whose passing Derrida observes are Jean-Francois Lyotard, Roland Barthes, Michel Foucault and Gilles Deleuze; and so, in his rapt attention to their work and their influence upon him, the book also offers a hesitant and tangential retelling of Derrida's own life in French philosophical history. There are illuminating and playful anecdotes—how Lyotard led Derrida to begin using a word-processor; how Paul de Man talked knowledgeably of jazz with Derrida's son. Anyone who still thinks that Derrida is a facetious punster will find such resentful prejudice unable to survive a reading of this beautiful work."—Steven Poole, *Guardian*

Paper $15.00

The Just

Paul Ricoeur

Translated by David Pellauer

"In essays on Rawls, Michael Walzer, and Ronald Dworkin, among American philosophers and Hannah Arendt and Luc Botanski among European theorists, Ricoeur sketches a distinctive position on questions of justice that straddles the liberal/communitarian divide common in the Anglo American world. . . . [Ricoeur provides] a fresh perspective on current debates within his own interesting account of the structure of moral action."—Georgia Warnke, *Ethics*

Paper $14.00

Thinking Biblically

Exegetical and Hermeneutical Studies

André LaCocque and Paul Ricoeur

Translated by David Pellauer

"This is two books under one cover, with two authors whose readings at times exhibit an exhilarating tension. In alternating chapters, LaCocque's exegeses take up the *how* and *who* of the interpretation, while Ricoeur explores the question of what allows for plurivocity in these particular texts. LaCocque's essays elegantly show the way in which passages are brought into new configurations by the readers and writers of the Bible."
—Paul-Jonathan E. Benson, *Boston Book Review*

Paper $18.00

The Moment of Complexity

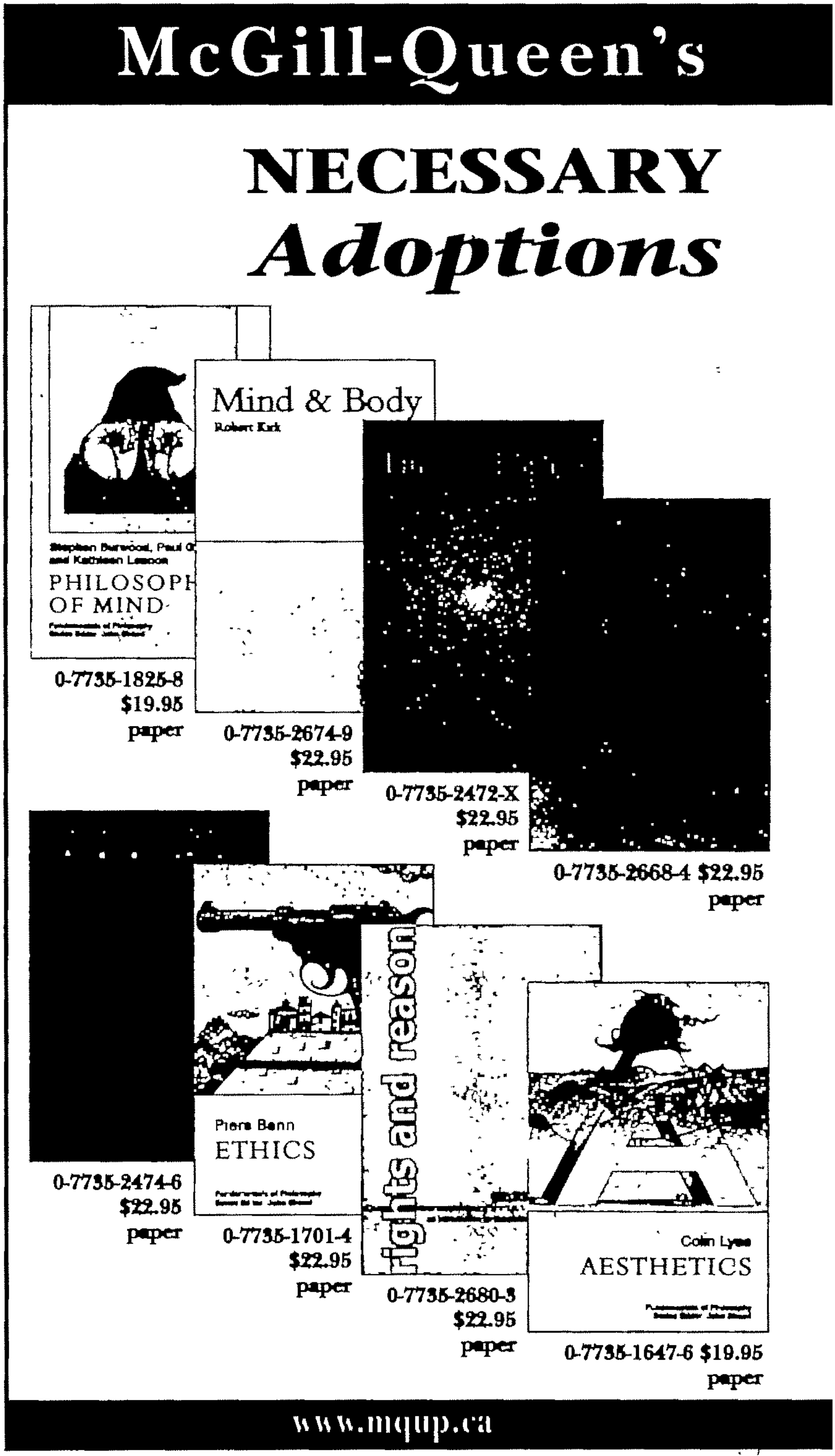
Emerging Network Culture

Mark C. Taylor

"This is a grand statement, written by somebody who is clearly frustrated by the marginalization of academic and intellectual discourse. It is also the staking of a claim, over the new sciences of complexity, as the location from which new paradigms of critical theory might emerge to supersede the moribund strategies of Postmodernism."—Charlie Gere, *Journal of Visual Culture*

Paper $17.00

The University of Chicago Press



McGill-Queen's

NECESSARY *Adoptions*

Stephen Burwood, Paul G
and Kathleen Lennon
PHILOSOPH
OF MIND
Fundamentals of Philosophy
Series Editor John Shand

0-7735-1825-8
$19.95
paper

Mind & Body
Robert Kirk

0-7735-2674-9
$22.95
paper

0-7735-2472-X
$22.95
paper

0-7735-2668-4 $22.95
paper

0-7735-2474-6
$22.95
paper

Piers Benn
ETHICS
Fundamentals of Philosophy
Series Editor John Shand

0-7735-1701-4
$22.95
paper

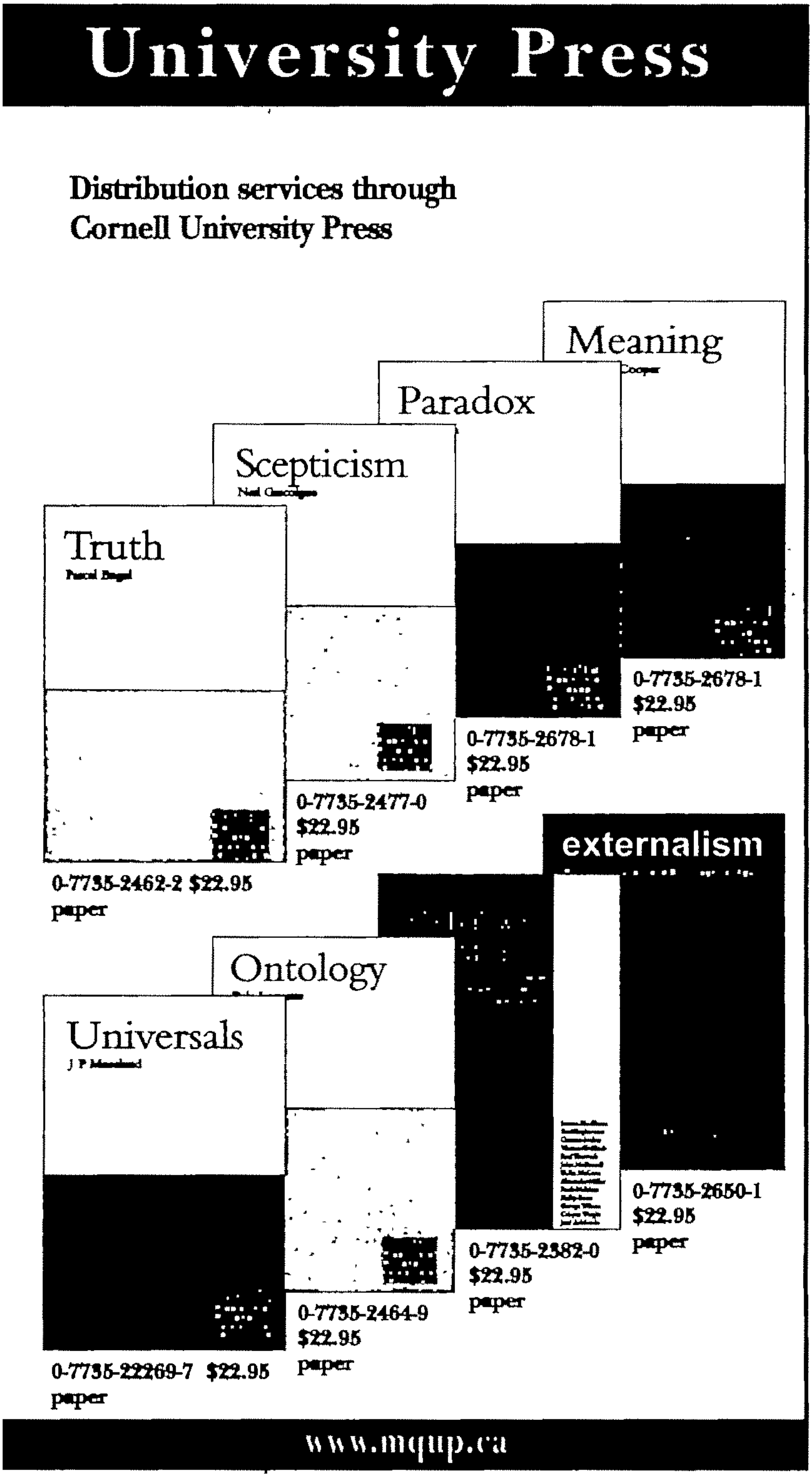
rights and reason

0-7735-2680-3
$22.95
paper

Colin Lyas
AESTHETICS
Fundamentals of Philosophy
Series Editor John Shand

0-7735-1647-6 $19.95
paper

www.mqup.ca



University Press

**Distribution services through
Cornell University Press**

**0-7735-2462-2 $22.95
paper**

**0-7735-2477-0
$22.95
paper**

**0-7735-2678-1
$22.95
paper**

**0-7735-2678-1
$22.95
paper**

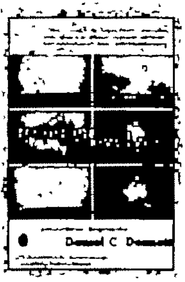
**0-7735-22269-7 $22.95
paper**

**0-7735-2464-9
$22.95
paper**

**0-7735-2382-0
$22.95
paper**

**0-7735-2650-1
$22.95
paper**

www.mqup.ca

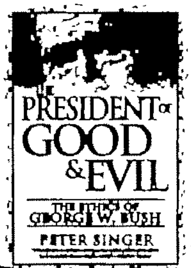


FREEDOM EVOLVES

Daniel C. Dennett

An "intellectually liberating" (*San Francisco Chronicle*) exploration of free will and morality—from the acclaimed author of *Darwin's Dangerous Idea*. "Crisp and critically insightful."—*The Washington Post Book World*

Penguin 368 pp 0-14-200384-0 $17.00



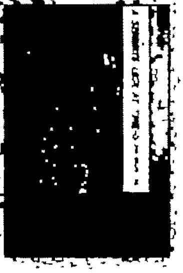
THE PRESIDENT OF GOOD AND EVIL

The Ethics of George W. Bush

Peter Singer

"Peter Singer may be the most controversial philosopher alive, he is certainly among the most influential."—*The New Yorker*

Dutton 288 pp 0-525-94813-9 $24.95



FASTER THAN THE SPEED OF LIGHT

The Story of Scientific Speculation

João Magueijo

"A new species of science book . . . provides the vicarious thrills that the poetically or mystically minded reader looks for in a book on theoretical physics."—*The Village Voice* "A racy account... jaunty and irreverent."—*The New York Times*.

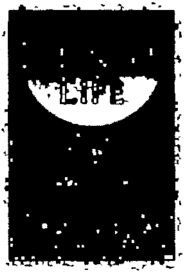
Penguin 288 pp 0-14-200361-1 $15.00

A SIDEWAYS LOOK AT TIME

Jay Griffiths

An "amusing and erudite" (*The Times Literary Supplement*) exploration of the way we experience time. "An irresistibly provocative and political analysis of time in our personal lives ... It could cause a revolution."—*The Times* (London), Books of the Year.

Tarcher 416 pp 1-58542-306-8 $15.95 April 2004



THE NICOMACHEAN ETHICS

Aristotle

Translated by J. A. K. Thomson

Revised with Notes and Appendices by Hugh Tredennick

Introduction and Bibliography by Jonathan Barnes

This authoritative edition features a lucid translation, new introduction, updated suggestions for further reading, and a chronology of Aristotle's life and works.

Penguin 384 pp 0-14-044949-3 $11.00

INFINITE LIFE

Seven Virtues for Living Well

Robert A. F. Thurman

Foreword by His Holiness the Dalai Lama

"Robert Thurman is a living treasure, one of today's most provocative spiritual thinkers."—Daniel Goleman.

Riverhead 288 pp 1-57322-267-4 $24.95

PENGUIN GROUP (USA)

Academic Marketing Department, 375 Hudson St., NY, NY 10014 www.penguin.com/academic

THE JOURNAL OF
PHILOSOPHY

VOLUME CI, NUMBER 2
FEBRUARY 2004

page
55 *Logic and Absolute Necessity* Scott A. Shalkowski
83 *Scanlon on Promissory Obligation: The Problem of Promisees' Rights* Margaret Gilbert

110 NOTES AND NEWS

[illegible]: 11\5\09

Published by **The Journal of Philosophy, Inc.**

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J E WOODBRIDGE AND WENDELL T BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Akeel Bilgrami, Arthur C. Danto, Kent Greenawalt, Patricia Kitcher, Philip Kitcher, Isaac Levi, Mary Mothersill, Philip Pettit, Carol Rovane, Achille C. Varzi. *Editor Emeritus*: Sidney Morgenbesser. *Consulting Editors*: David Albert, John Collins, James T. Higginbotham, Charles D. Parsons, Wilfried Sieg. *Managing Editor*: John Smylie.

THE JOURNAL OF PHILOSOPHY is owned and published by the Journal of Philosophy, Inc. *President*, Arthur C. Danto; *Vice President*, Akeel Bilgrami; *Secretary*, Daniel Shapiro; *Treasurer*, Barbara Gimbel; *Other Trustees*: Lee Bollinger, Leigh S. Cauman, Kent Greenawalt, Michael J. Mooney, Lynn Nesbit.

All communications to the Editors and Trustees and all manuscripts may be sent to John Smylie, Managing Editor, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, New York 10027 FAX: (212) 932-3721. You may also visit our website at: www.journalofphilosophy.org

THE JOURNAL OF PHILOSOPHY

2004

SUBSCRIPTIONS (12 issues)

Individuals	$35.00
Libraries and Institutions	$75.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$15.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are required.

Published monthly as of January 1977; typeset and printed by Capital City Press, Montpelier, VT.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL's terms for all copying beyond that permitted by Sections 107 or 108 of the U S Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

POSTMASTER Periodical postage paid at New York, NY, and other mailing offices.

POSTMASTER. Send address changes to the *Journal of Philosophy* at MC 4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027.

© Copyright 2004 by the Journal of Philosophy, Inc. ISSN 0022-362X

THE JOURNAL OF PHILOSOPHY
VOLUME CI, NO. 2, FEBRUARY 2004

LOGIC AND ABSOLUTE NECESSITY*

> Philosophy, including Logic, is not primarily about language, but about the real world.
> —A. N. Prior

It is hard to conduct our conceptual lives without some recourse to modality. Necessary and sufficient conditions are the stuff of conceptual analysis, laws of nature and nomic determination inform the philosophy of science as well as the philosophy of religion, and mathematical truth seems a poor candidate for arbitrariness and convention. Even the most squeamish about metaphysics in general and modality in particular will find that, in at least a range of cases, a proposition entails its own possibility, making it difficult to eschew completely all modal discourse.

The main issue of contention, of course, is not whether we self-consciously avail ourselves of modal idioms but the substance behind those idioms. Logical necessity enjoys a rather privileged status in the minds of most analytic philosophers, largely because it is thought to be an innocent modality. Even W.V. Quine is willing to countenance logical necessity to the degree that it commits him to nothing more than logical truth.[1] The developments of formal logic, especially model-theoretic semantics, have paved the way for apparently clear accounts of the relation of logical consequence and the correlated property of being a logical truth. With one or other of these in hand, logical necessity is easily definable: A is logically necessary if and only if A is a logical truth if and only if A is a logical consequence of the axioms of logic. With logical necessity thus understood in formally

---

* I am grateful to those who commented on presentations of these ideas at the Universities of Arizona, Leeds, Liverpool, and Notre Dame and to Otavio Bueno, John Divers, and Joseph Melia, for comments on an earlier version of this paper.

[1] "Comment on Marcus," in Robert B. Barett and Roger F. Gibson, eds., *Perspectives on Quine* (New York: Blackwell, 1990), p 244.

0022-362X/04/0102/55–82

manageable terms, other metaphysically significant necessities, such as physical necessity, can be defined as the logical consequences of some specified set of nonlogical axioms. The physical necessities would be, in this framework, the logical consequences of the fundamental laws of physics. If logical necessity is admitted as the most basic necessity, we can make sense of contingent necessities—those for which the basic axioms are contingent, as the laws of physics are thought to be. There is, thus, a great attraction to treating logical necessity as the most basic necessity.

While some residual bush fires regarding the proper formalization of logical necessity may require some attention, there would be nary a philosophical scruple against modality were all metaphysically significant modalities definable in the above fashion. It is the revival of Aristotelian essentialism and the invocation of possibilia that has caused the contemporary storm of controversy. Essentialists maintain that, unlike physical necessities, metaphysical necessities are necessary *tout court* and not merely derivatively, and both essences and possibilia seem hopelessly entangled in mysteries not amenable to empirical resolution. Granting the necessity of logic seems safe and free of epistemologically troublesome metaphysical commitments. This is the standard intuitive basis for admitting logical necessity and closing the door firmly to any modality not open to treatment in terms of logical notions. For my purposes, this form of selective modalizing constitutes anti-essentialism.

In what follows I argue that this received view about modalities is seriously misguided. The most general and fundamental necessity does not rest on logic and the semantics of the logical constants but rather on the natures of objects, where the objects involved are ordinary concrete objects and their properties and/or propositions and their properties. Logical necessity is best understood as a special case of metaphysical necessity. Furthermore, I will argue that disagreements over the precise formalization of logical necessity, as well as the viability of essentialism, are, in reality, disagreements over particular substantial metaphysical theses, thus showing that even the philosophy of logic is itself entangled with "metaphysical speculation."

In section I, I present Bob Hale's argument that logical necessity is absolute necessity. I will argue that this argument fails to deliver the priority of logical necessity as traditionally construed. Sections II and III contain brief discussions of psychologism and formalism and a lengthier treatment of the model-theoretic approach to the nature of logic. The former two approaches provide nothing like necessity while the latter does so only insofar as it presupposes some quite distinct necessity. The moral is that model theory, as it is typically understood, cannot do it all on its own. In section IV, I argue that contemporary accounts of logical consequence are no help to the anti-essentialist even though they add a modal component to the Tarksian model-theoretic account of consequence and, in section v, I sketch an essentialist account of logical necessity that promises to accommodate ordinary and philosophical argumentative practices in a way that standard anti-essentialist accounts do not.

I. HALE'S DEFENSE

While there are obvious benefits in holding logical necessity to be the most fundamental or basic necessity, there are surprisingly few attempts to defend such a position. Most take the form of rear-guard defenses that accuse other approaches to modality of suffering from defects too great to tolerate, usually of the epistemological variety. The trouble with this general strategy is that defects, like beauty, are often in the eye of the beholder. As a notable exception, Hale, taking inspiration from Ian McFetridge, defends the primacy of logical necessity to this extent: he argues that there is no sense in which a logical necessity could be false.[2] At this stage we may treat this as a necessary condition of a basic necessity.

In comparing two different kinds of necessity Hale stipulates that one is *at least as strong as* the other just in case all necessities of the one type are necessities of the other type, leaving open the question of the reverse inclusion. A kind of necessity is *absolute necessity* just in case it is at least as strong as every other kind of necessity (*op. cit.*, pp. 94–95). It is this kind of absoluteness that Hale claims for (broadly) logical necessities—necessities which are either straightforward logical truths or else are analytic or otherwise conceptual truths. If logical necessity thus understood is absolute, then all others are relative necessities in the same way physical necessities are relative to logical necessities—they are consequences of some nonlogical axioms. The set of logical necessities is the most restrictive set of necessities, one included in all others. This, however, does not seem to square with the posture of most essentialists. Typically, essentialists do not think that essentialist claims are merely relative necessities, since they think that essentialist theses are both necessary and not fully definable in terms of logical necessity. In effect, essentialists treat metaphysical necessity, not logical necessity, as the most fundamental.[3]

---

[2] Hale, "Absolute Necessities," *Philosophical Perspectives*, x (1996): 93–117; McFetridge, "Logical Necessity: Some Issues," in John Haldane and Roger Scruton, eds., *Logical Necessity and Other Essays* (London: Aristotelian Society, 1990), pp. 135–54. Hale understands 'logical necessity' broadly, to encompass logical truths as well as conceptual necessities. He says that austerely logical necessity, the necessity associated with only with logical truths, is a special case of broadly logical necessity It is easy to see what he has in mind when we think of conceptual necessities as those true (necessarily) in virtue of the meanings of the relevant constituent expressions. Austerely logical truths are necessarily true in virtue of the constituent logical concepts. Broadly logical truths are necessarily true in virtue of constituent nonlogical concepts as well as the relevant logical concepts. On the assumption that the class of broadly logical (conceptual) necessities includes all of the austere logical truths, my discussion focuses on austerely logical truths/necessities. If the case for the absoluteness of austere logical necessities fails, then the case for broadly logical necessity fails

For his direct defense of the absoluteness of logical necessity, Hale explicitly makes five assumptions, where '$\Box_L$' denotes logical necessity and '$\Diamond$' denotes an arbitrary kind of metaphysically significant possibility such that $\Box A \rightarrow A$.

(A1) If $\Box_L(A \rightarrow B)$ then $\Box_L((A \,\&\, C) \rightarrow B)$
(A2) $\Box_L(A \rightarrow A)$
(A3) If $\Box_L(A \rightarrow B)$ and $\Box_L(A \rightarrow C)$ then $\Box_L(A \rightarrow (B \,\&\, C))$
(A4) If $\Diamond A$ and $\Box_L(A \rightarrow B)$ then $\Diamond B$
(A5) $\neg\Diamond(A \,\&\, \neg A)$

(A1)–(A3) are standard assumptions about logical consequence, (A4) is a typical assumption about deductive closure for metaphysically significant modalities, and (A5) merely formalizes the assumption that contradictions are not possibly true.

The argument is the following reductio.

(1) $\Box_L(A \rightarrow B)$	Assumption
(2) $\Diamond(A \,\&\, \neg B)$	Assumption
(3) $\Box_L((A \,\&\, \neg B) \rightarrow B)$	from (1), by (A1)
(4) $\Box_L(\neg B \rightarrow \neg B)$	(A2)
(5) $\Box_L((A \,\&\, \neg B) \rightarrow \neg B)$	from (4), by (A1)
(6) $\Box_L((A \,\&\, \neg B) \rightarrow (B \,\&\, \neg B))$	from (3), (5), by (A3)
(7) $\Diamond(B \,\&\, \neg B)$	from (2), (6), by (A4)
(8) $\neg\Diamond(B \,\&\, \neg B)$	(A5)
(9) $\neg\Diamond(A \,\&\, \neg B)$	from (2), (7), (8), by reductio

---

[3] This is true of most essentialists, though in "Essence and Modality," *Philosophical Perspectives*, VIII (1994): 1–16, and "Ontological Dependence," *Proceedings of the Aristotelian Society*, XCV (1995): 269–89, Kit Fine distinguishes between essence and metaphysical necessity. Think of metaphysical necessities as those given by what is true in all possible worlds (or possible worlds in which the relevant objects exist). Essentialist necessity, to give it a distinct label, is more fine-grained. It might be metaphysically necessary that Socrates be a member of singleton Socrates or that all equiangular triangles are equilateral. In neither case, though, do these facts about what obtains in all possible worlds provide a truth about the essence or nature of either Socrates or equiangular triangles. Essentialist necessities form a subset of metaphysical necessities, a subset that expresses the nature or Aristotelian real definition of the relevant items. Because nothing in Hale's argument or my discussion depends on this distinction, I will make nothing of it in what follows.

The main issues that must be confronted when assessing this argument are: How is '$\Box_L$' properly defined and how do the argument's assumptions and premises bear on essentialism? First, to the proper scope of the logically necessary. Hale distinguishes between *austerely* and *broadly* logical necessities. Austerely logical necessity encompasses only the first-order logical truths—those depending only upon the meanings of the logical constants for their truth. Broadly logical or conceptual necessity encompasses the addition of necessities that depend upon the meaning of nonlogical concepts. So, it is not an austerely logical necessity that all vixens are female, but it is a broadly logical necessity. It is an austerely logical possibility that there be male vixens, but it is not a broadly logical possibility. Hale's preference is to understand '$\Box_L$' as broadly logical necessity. This preference, though, undermines the claim that '$\Box_L$' expresses absolute necessity. By the definition of absoluteness, austerely logical necessity is stronger than broadly logical necessity, so broadly logical necessity is not absolute necessity. In fact it is a rather trivial exercise to generate stronger and stronger necessities. By progressively reducing the axioms used to define a necessity, progressively stronger necessities emerge. One that is very familiar is propositional necessity—the modality that admits as necessary only the truths of propositional logic. Given the multiplicity of logics, there is no one necessity that, from a strictly formal point of view, deserves the honorific title of "logical necessity." Considerations beyond this reductio must be adduced to justify any specific choice of logics to serve as the basis for defining '$\Box_L$' even when we limit our considerations to traditional logics. Thus, even if all of the assumptions of the argument are permitted to stand, the argument is inadequate to establish the absoluteness of Hale's preferred modality.

The assumptions should not be granted so easily, however. There are necessities definable from very nontraditional logics. In some well-developed paraconsistent logics, the Law of Noncontradiction fails and so, according to (logical) necessities specified in accordance with the truths of these logics, it *is* possible that there be true contradictions. Thus, (A5) is, strictly speaking, false and will not be granted by all anti-essentialists. There are senses of 'possible' according to which it is possible that for some A, $A \mathbin{\&} \neg A$. There may well be ways of justifying a preference for traditional logics and their necessities over paraconsistent logics and their necessities. It might be argued that one can make nothing of contradictions or situations in which contradictory propositions are true. This would, at least, be a relevant ground for choosing one logic as a correct logic. The point remains that this argument, with its assumed principles, is insufficient to establish the absoluteness of broadly (classical) logical necessity.[4]

In order to discount the significance of these logics in this context, Hale himself must justify the claim that any necessity that is stronger than broadly logical necessity—like austerely logical, propositional, or a paraconsistent necessity—is one that can be safely ignored. One basis for ignoring them would be that they are merely formal and not genuine necessities, where a merely formal necessity is one that is clearly defined but also one that does not fully characterize the relevant necessary truths. To put it in terms of possibility, Hale must defend the claim that any necessity that is stronger than broadly logical necessity admits as possible things that are not genuinely possible; it admits as possibly true propositions that are not possibly true. He clearly does this in his attempt to set aside austerely logical necessity (*op. cit.*, pp. 100–01). He argues that while one can *consistently* claim that there are male vixens, there is no possible world in which there are male vixens. This would be a quite trivial observation to make if this remark were intended to mean only that there are no *broadly logical* possible worlds in which there are male vixens. All parties can agree to this. The point is rescued from triviality only if the existence and nonexistence of certain kinds of possible worlds is supposed to have something to do with *genuine* possibility, with the real truths about possibility. Hale's claim, though, is unlikely to have much dialectical force against one who claims that austerely logical necessity is the absolutely genuine necessity, since such a one will think that there *are* possible worlds, if they think in terms of possible worlds at all, in which there are male vixens. Or, if the discourse of possible worlds is shunned, such a one will simply affirm that male vixens are genuinely possible. The argument under consideration, then, cannot be completely general in establishing the absoluteness of broadly logical necessity. It must be supplemented with considerations about the scope of the genuinely possible if it is to avoid begging important questions and for some this issue is bound up with which logic is the correct logic and the meaning of '$\Box_L$'. So much for issues bearing on the scope of '$\Box_L$'. On to the relevance of the argument to essentialism.

For traditionally minded essentialists, that Hale's argument is insufficiently general is somewhat cold comfort. It does not beg the question against them after all, since they accept (A5) and many concede that '$\Box_L$' encompasses all and only the broadly logical truths. Accepting the other assumptions forces essentialists to look more carefully at the argument itself, which is a traditionally valid reductio. A close examination of the premises, however, shows that neither (1) nor (2) is a consequence of essentialist claims. Such claims usually involve logical possibilities that are not metaphysical possibilities; they do not involve metaphysical possibilities that are not even logical possibilities. An essentialist might claim that while it is logically possible that Kripke's lectern have been made of ice, it is not really possible for it to have been so. There is no contradiction in supposing that Queen Elizabeth II might have had different parents, even though certain facts about her parentage are essential to her. Thus, essentialist claims really amount to the following variations on (1) and (2) where '$\Box_M$' is an operator expressing metaphysical necessity:

---

[4] For a development of a theory in which contradictions are sometimes both logically possible and genuinely possible, cf. Graham Priest, *In Contradiction* (Boston: Kluwer, 1987).

(10) $\Box_M(A \rightarrow B)$
(11) $\Diamond_L(A \mathbin{\&} \neg B)$

Thus, metaphysical possibility is supposed by essentialists to be *less* permissive, not more permissive, than logical possibility. Accordingly, essentialists hold that logical possibility, as given by first-order quantificational logic with identity and perhaps any relevant nonlogical meaning postulates, overgenerates admissible formulae and counts as possible things that are not really possible.

Now the core of the dispute between essentialists and their critics cannot sensibly be construed as a dispute over the scope of '$\Box_L$' or '$\Box_M$'. While there may be disputes over the scope of the logically necessary (as we have just seen) and while there may be disputes over which things are not logically necessary but are metaphysically necessary, these are secondary disputes. The fundamental dispute is surely over which, if either, of these modalities correctly characterizes the necessary *truths*. It is a dispute over whether everything not ruled out as false by the relevant logical truths is a genuine possibility, which just is the dispute over whether there are more necessary truths than the relevant logical truths. Thus, the reductio stemming from (1) and (2) serves to refute precisely the exact opposite of what is relevant to rejecting essentialism. (2) is the supposition of the falsehood of the logically necessary truth postulated in (1). The reductio is really a refutation of the thesis that first-order logic *under*generates possibilities—that logical possibility is a reliable guide to real possibility so far as it goes; it just does not go far enough. But, typical essentialists have had no interest whatsoever in denying any logical truths and diminishing the scope of the necessary. They are in the business of expanding the scope of the necessary by adding nonlogical metaphysical necessities. Thus, the logically traditional essentialist should maintain that Hale's argument successfully shows that (1) and (2) are indeed inconsistent, but that no pairs of propositions of those forms are entailed by essentialism. Essentialist claims have the form of (10) and (11), but the argument does not even so much as address them. Essentialists can readily grant that no real possibilities outstrip the broadly logical possibilities because they do not even grant that all of the broadly logical possibilities are genuine possibilities, so they certainly should not hold that some genuine possibilities are not even logical possibilities. Thus, this argument does nothing to undermine essentialism.

Hale thinks that the essentialist should be uncomfortable with the admission that something might be metaphysically necessary but not logically necessary because metaphysical necessity would not, then, be absolute necessity. "To accept that metaphysical necessity is not absolute is to acknowledge that while it is, say, metaphysically necessary that heat is mean kinetic energy of molecules, there are possible worlds—logically possible worlds—in which this is not so" (*op. cit.*, p. 98).

This is no real problem for the essentialist, however. 'Absolute necessity' is a term of art and its application to metaphysical necessity is not something to which the essentialist is wedded. That there are more restrictive necessities that are merely formally characterized necessities is of no consequence to the essentialist, who is concerned only to maintain that for any more restrictive necessity the correlated kind of possibility is too permissive and admits some counterfeit possibilities. The mere fact that a necessity is more restrictive is, by itself, no reason to think that it retains a proper purchase on modal reality.

Furthermore, given the existence of necessities more restrictive than broadly logical necessity, Hale is in precisely the same dialectical predicament as the one in which he wishes to put the essentialist. If the essentialist must be discomforted by the admission that there are logically possible worlds in which characteristic essentialist claims are false, Hale must likewise be discomforted by the admission that there are austerely logically possible worlds in which characteristic broadly logical truths are false. Likewise for propositionally- and paraconsistently-possible worlds.

Of course the essentialist just denies that there are any merely logically possible worlds, if they choose to treat possible worlds as real existences, and Hale just denies the existence of the austerely logically possible worlds. Even if we set aside worries over paraconsistency and grant the epistemological security and metaphysical innocence of some traditional set of first-order logical truths, both Hale and the essentialist must justify their favored restriction on the set of austerely logical possibilities in order to obtain precisely their respective chosen sets of genuine possibilities. One who makes sense of necessity only insofar as it is coextensive with first-order logical truth will want reasons for placing any restriction at all on the class of austerely logical possibilities, whether this restriction is given in terms of the meanings of nonlogical terms or in terms of essences. While the essentialist must justify the verdict that some consistent sentences which violate "metaphysical postulates" fail to express real possibilities, the conceptualist must justify the verdict that some consistent sentences that violate meaning postulates fail to express real possibilities. No advantage to the conceptualist here.

Furthermore, Hale has no particular epistemological advantage in his appeal to the a priori knowledge of concepts. Typical contemporary essentialism derives from allegedly a priori principles regarding identity, chemical constitution, origins, and the like. The reasoning that serves to justify these principles is no different in character from the kind of conceptual reasoning that would be used to justify the impossibility of male vixens. Whether we choose to call the essentialist's a priori principles conceptual (or semantic) principles depends solely on whether we are externalists or internalists about meaning. The essentialist might depend upon empirical investigation to provide the fine-grained details of necessities about Kripke's lectern or the Queen, but these bits of nonconceptual knowledge are completely uncontroversial. The most important, and most difficult, work for both the essentialist and the conceptualist is a priori. Each must defend the a priori status of their chosen principles against the Quinean who has no time for distinctively analytic or essentialist truths. Again, no advantage to Hale.

Thus, Hale's preferred position affords no dialectical advantage over the essentialist. Each must stray from the apparent security and innocence of austerely logical truth; each must justify their respective chosen limit on genuine possibility; each derives necessities largely from a priori investigation, with perhaps the addition of merely uncontroversial empirical information. Thus, Hale's argument fails to establish the absoluteness of logical necessity, traditionally understood. It is vulnerable to criticism from those who are more profligate in admitting possibilities and it is irrelevant to the essentialist who is more restrictive. Furthermore, Hale's position suffers from difficulties parallel to those of the essentialist. Similar parallels will be noted below.

II. LOOKING AT LOGIC

So far we have only one unsuccessful argument for the priority of the logical modalities. To forestall the search for a better argument for this thesis, I begin an extended argument for a contrary thesis—that metaphysical necessity is prior to logical necessity—by examining the nature of logic. The nature of logic is the subject for a rather wide-ranging discussion, one that must address the relation between modern mathematical logic and everyday inference, the way logical knowledge can be used as a tool to extend our nonlogical knowledge, the nature and significance of logical form and the relation between "logical facts" and ordinary states of affairs—a subject well beyond the scope of this discussion. I will limit myself to considerations of (austerely) logical truth and logical consequence, since one or other of these notions provides the substance of logical necessity and is least subject to philosophical controversy.

Logical truth and consequence, arguably, can be illuminated by reference to psychological, syntactic, and semantic facts.[5] If an account of the nature of logic is to yield a treatment of something that can be properly dignified as a kind of genuine necessity, the psychological and syntactic elements of logic are inadequate. Very brief reasons for this are the following.

Psychologism about logic must be given in either wholly descriptive or partially normative terms. Descriptive psychologism yields logic as a subject of empirical, not a priori, study. If, on the one hand, relations of logical consequence are the laws of thought and these are universal patterns of human thought, then the account is too strong and entails that there are far fewer relations of consequence (and correlated logical truths) than standard logical theory recognizes; we sometimes fail to see the consequences of our beliefs. If, on the other hand, the laws of thought are those patterns used by someone or other at some time or other, then the account is too weak and entails that there are far more relations of consequence and more logical truths than typically recognized. Sometimes we reason invalidly. Normative psychologism encounters similar problems. If logical consequence were to be what should be believed by a rational agent given other beliefs, then we must relinquish either logical theory as it currently stands or the belief that some of us are sometimes rational. Any belief has infinitely many consequences, according to standard logical theories, yet apparently rational agents fail to believe all of these consequences.

---

[5] It is noteworthy that Quine's own treatment of logical truth is an admixture of all three approaches to logic. In *Methods of Logic* (New York: Holt, Reinhart, Winston, 1972), p. 4, he explains logical consequence in psychological terms, in *Philosophy of Logic* (Englewood Cliffs, NJ: Prentice-Hall, 1970), pp. 47–60, he provides several different accounts that he takes to be coextensive.

A naturalistic account of rational inference makes logical revisionism almost inevitable.

Formalism fairs no better than psychologism. Formal facts on their own have nothing whatever to do with truth, much less logical truth and logical consequence. Syntactic specifications of the elements of a language, along with formation and derivation rules, constitute nothing more than the delineation of a game. Here are the pieces; this is how to build things with them; that is how to move them about. Devoid of interpretation, we can go no further. What can be "proved" on the basis of other things is completely uninteresting without some substantial constraints on syntactic form and the rules of derivation. For example, we sometimes think of logical truths as those formulae that can be proved from the null class of assumptions. With no constraints on the rules, however, we are always free to introduce whatever rules are necessary to prove whatever we like from the null class. This seems preposterous, but only because in practice the introduction of "rules of inference" is constrained by semantic information.

For these and other reasons, the semantic approach to logic has been given priority as exposing the most important elements of the nature of logic. In what remains I shall argue not that model-theoretic accounts of metalogical notions are extensionally incorrect, but rather that model-theoretic accounts cannot work to secure the priority of logical necessity as the proper base level necessity. The model-theoretic accounts succeed only insofar as they either tacitly embed some more basic necessity or else anti-essentialists are not entitled to make their broad assertions about the role that model theory can play in modal theory.

III. WHAT MODEL THEORY CANNOT DO FOR YOU

Given standard soundness and completeness results for first-order logic, we know that for a set of formulae, Γ, and any formula, α, $\Gamma \nvdash \alpha$ if and only if $\Gamma \nvDash \alpha$, that α is provable from Γ using familiar rules if and only if all models in which all members of Γ are true are also models in which α is true. Furthermore, what can be proven from no assumptions using those rules is exactly what is true in all models; $\nvdash \alpha$ if and only if $\nvDash \alpha$. If provability does not expose the underlying nature of logic and logical necessity, then these metatheoretical results point to semantic facts as the most significant facts of logic. Thus, if the logical necessities are simply the logical truths, then logical necessity is truth in all models and logical possibility is truth in at least one model.

Of course, this looks very much like David Lewis's reductive account of modality.[6] Modality is traded for quantification, only in Lewis's account the domain of quantification is the domain of possible worlds, realistically and concretely construed. Initially, Lewis was greeted mainly with the infamous incredulous stare. Given the formal similarities between the nearly universally accepted account of logical truth and logical necessity and Lewis's account of necessity, the incredulity should not have been due to the mere trade of modality for quantification. It had to be the trade of modality for *possibilist* quantification that caused incredulity. It was Lewis's ontology that was troubling, not the structure of his account of necessity. On pain of inconsistency, then, the 'all' in the anti-essentialist metalanguage in which the concepts of logical truth and necessity are given must be an actualist quantifier. Quantification over models and their constituents is, thus, supposed to be innocent in a way that quantification over possible worlds and their constituents is not. Furthermore, the models themselves are supposed to be so innocent that it is difficult to find much attention given to their nature, aside from the observation that they consist in nonempty sets with relations and functions. I will canvass the important options.

For the extremely metaphysically wary, the nominalist who eschews abstracta of all kinds, the only actual models are concrete models. '$\Box_{\mathrm{L}}$' is defined over only the actual arrangements of concrete objects. The nominalist actualist recognizes nothing else. This stance overgenerates necessities and undergenerates possibilities. It overgenerates necessities because everything true in all actual concrete models is deemed logically necessary even when, on the normal face of it, things could have gone differently and some of those sentences true in all actual concrete models would have been false had things indeed gone otherwise. It undergenerates possibilities because this type of nominalism countenances nothing beyond actual concrete structures and, intuitively, the actual concrete world does not exhaust all possible structures. These mismatches between our intuitive modal judgments and the nominalist's definition of '$\Box_{\mathrm{L}}$' aside, one straightforward consequence of this account is that '$\Box_{\mathrm{L}}$' is undefined over contrafactual situations.[7] Logic, on this construal, has no application at all to anything beyond actuality. Surely, one constraint on any account of the nature of logic is that it explain how logic is a useful tool for reasoning about contrafactuality. So, this particular model-theoretic approach fails to capture traditional logics and their correlated modality.

---

[6] *On the Plurality of Worlds* (New York: Blackwell, 1986).

[7] I choose 'contrafactual' to be more general than 'counterfactual'. The former concerns any nonactual possibility, the latter only those relevant to the assessment of counterfactual conditionals, which are in turn closely aligned with pragmatic and perhaps causal considerations. Typically, those situations or worlds relevant to the truth value of a counterfactual conditional are a subset of the totality of actual and merely possible situations or worlds.

For a standard platonist alternative some abstract entities must be countenanced. To solve the problem of too few models and the undergeneration of possibilities, platonist model theory may be permitted to invoke a set-theoretic hierarchy that is as rich as one likes. Nothing in my further discussion relies upon any assumption about the richness of this hierarchy. There seems to be no obstacle to having enough objects to quantify over, so leaving out possibilities cannot be a problem to do with the size of the recognized ontology.

Platonism still faces the main affliction of nominalism, however. The recognized ontology remains actualist and nonmodal. The platonist strategy is simply to recognize a new class of objects in our ontology. Merely admitting more objects to the domain of quantification, however, does not by itself solve the problem of capturing the substance of logical necessity. The net result is the same for the platonist as it is for the nominalist: α is supposed to be a logical necessity just in case it is true in all members of the class of actual models. The platonist still leaves '$\Box_L$' undefined over contrafactual situations. So far, all the platonist has told us is that the actual world contains some abstract objects in addition to the concrete objects recognized by the nominalist. With respect to understanding modality this is no more illuminating than being told that there are extraterrestrial objects such as planets and comets in addition to the terrestrial objects we all know and love.

That abstract objects are noncausal or not spatiotemporally located presents no solution to the nominalist's problem unless immutability and necessity are conflated. The definition of abstract objects as those that enter into no causal relations entails only that no object causally depends for its state upon abstract objects and no abstract object causally depends for its state upon any other object. At most, the abstractness of models secures the *actual* and perhaps causal stability of the model-theoretic hierarchy, but not its modal stability.[8] When all is said and done, standard platonism delivers no more than the claim that there is a given set of actual models, some realized concretely, some not. What it does not directly address is whether the abstract models, unlike the nominalist's concrete models, cover all the possibilities.

---

[8] Whether even the actual stability of the hierarchy is guaranteed depends on whether uncaused changes are possible, which I will not discuss here.

At this point it is open to the anti-essentialist to find a disanalogy between the current account of logical necessity and Lewis's nonactualist account. Whereas for Lewis the operative notion was truth *in* a world, the model-theoretic approach to logical necessity depends on truth *according to* a model. '$\Box_L$' is defined appropriately not because it is defined over a class of nonactual objects but because it is defined over a class of actual objects that *represent* both actual and nonactual states of affairs. The actualist ontology "speaks" about both actual and nonactual realities.[9] So '$\Box_L$' is defined appropriately because it is defined over the possible as it is represented in the model-theoretic hierarchy. Thus, it is assumed that (i) every model represents a possibility and that (ii) all possibilities are represented by some model. These added representational constraints take us to modal platonism, which solves the problem of standard platonism by making explicit that what does the crucial work in the proposed account of logical necessity is not truth-in-a-model, but truth-according-to-a-model. It is the representative function of models that is important, not merely their existence and internal structure. Even though the ontology remains actualist there is some interesting connection between this class of actual objects and possibilities. '$\Box_L$' is now finally defined over objects which are themselves defined on real possibilities. Ordinary contrafactual reasoning is rescued.

That the first constraint is required is shown by the method of counterexample. An argument is shown to be fallacious by "finding" a model in which the premises are true, but the conclusion false. Were the selected model to represent an impossibility, then it would be incorrect, on that basis alone, to conclude that the argument under scrutiny is fallacious. We simply assume that anything we countenance as a model embodies or represents a possibility. The failure of the set of models to satisfy the second constraint guarantees that some α will be deemed a necessity when it is not. It would be false in one of the possibilities unaccounted for by the set of recognized models.

Modal platonism provides the basis for the most plausible model-theoretic accounts of the logical modalities. Is there any nonquestion-begging basis for thinking this account successful, however? Determining the answer depends very much on what the intent of the account is. If the anti-essentialist proposes it in the interests of conceptual or meaning analysis, then the account is a clear failure. A successful analysis of a modality must be given either in terms of some quite distinct modality or in wholly nonmodal terms and the analysis must be semantically equivalent to what is analyzed. The anti-essentialist maintains that the logical modality is the most basic of the relevant metaphysical modalities, so the first option is unavailable. Though it is hard to consult thoroughly untutored semantic intuitions, it is doubtful that what we meant by 'α is a logical truth' is that α is true according to all platonistically-acceptable models. If the account stops there, then not only is it doubtful that it really captures what was meant prior to the popularity of the model-theoretic account but precisely because of this there is no reason to think that the account is even extensionally adequate. That α is true according to all platonistically-acceptable models does not guarantee that all possibilities are represented nor that no inconsistencies are represented. Of course, conditions (i) and (ii) can be used to secure extensional adequacy, but at the cost of making the account circular, since the meaning of the relevant portion of modal discourse is given in terms of discourse about models and the proper meaning of talk about models is given, ultimately, in the modal terms of (i) and (ii). So, as a piece of conceptual analysis, the model-theoretic account fails.

---

[9] That a structure can represent something independently of any acts of interpretation by cognizers is a very dubious claim, but I shall not challenge it here.

One keen on the model-theoretic proposal can certainly make the case that there is no clear fact about what we meant by 'α is a logical truth' prior to the development of model theory. Perhaps our talk about logical truth, logical necessity, and logical possibility was vague, unruly, and not properly understood. If so, then it is not best to view claims about logical truths as semantically equivalent to their model-theoretic substitute. Nevertheless, its proponents will maintain, the model-theoretic account does the best job of systematizing and clarifying our somewhat ill-behaved usage. The point is not semantic analysis, but metaphysical equivalence or reduction. The anti-essentialist asks: Is it not the case that something is countenanced as a logical truth if and only if it is true according to all platonistically-acceptable models? If so, job done. Forget about whether we can define the notions in more basic conceptual terms. As metaphysicians we are after the ultimate structure of reality, not the most basic conceptual building blocks. So long as the relevant biconditional is true the theory should be deemed a success and since it takes things we find puzzling (modalities) and shows that they are interestingly related to things we understand better (quantification and sets), then it constitutes a philosophical advance.

An essentialist, however, can press three related problems against this version of the modal platonist. The first, which I shall not press here because it is well known, is the obvious analogue of the objection raised by Paul Benacerraf, and extended by Hartry Field, against mathematical platonism—the apparent epistemic inaccessibility of abstract objects.[10] The second problem is somewhat peculiar to the present anti-essentialist. In the present context, is the anti-essentialist entitled to maintain that something is a logical truth if and only if it is true according to all models? What is the basis of the modal platonist's confidence that abstract models satisfy the two modal constraints, that is, that there is a sufficiently close match between our considered modal judgments and the structure of the set-theoretic hierarchy? Sometimes, when speaking of models, the act of construction is given prominence.[11] Suppose, contrary to the typical intent of model theorists, that constructing a model is bringing that model into being, rather than showing that there is such a model. Given standard assumptions about the concrete and finite nature of acts of construction, constructed models are not the right sort of thing for the logically orthodox platonist who stands against nominalism. There are too few actually constructed models to be matched with what we antecedently take to be all the logical possibilities. Construction obviously aids with logical epistemology while undercutting the motivation for the move away from nominalism to platonism. For this reason platonists are more inclined to adopt an epistemology that makes the doing of formal logical semantics and mathematics a process of discovery rather than construction. There is nothing, though, in standard model theory that can count as justifying the two modal constraints that modal platonism requires. It is not as though model theorists have exhaustively examined objectively existing entities, catalogued their structures and their representational capacities, compared the results of this examination with considered judgments about the modal facts, and then reported back to the rest of us. Neither have they done a reasonable "empirical" examination of some models, formed inductive hypotheses, tested the hypotheses and then, when there was sufficient justification, gone public with their results.

The epistemology of model theory relies upon what mathematicians and logicians have antecedently taken to be valid inference as well as what they have taken to be initially plausible candidates for axioms of set theory. Determining what models there are, for example, determining whether there can be countable models for theories of certain sorts, assumes facts of logical consequence and logical truths. For instance, it is assumed by classical logicians that models obey the standard principles of classical logic: Noncontradiction and Excluded Middle. When we arrive at surprising results like the Löwenheim-Skolem theorems, our use of assumed logical facts do all the work in telling us about the number and nature of models. Whether the process is one of construction or discovery, the logical dog is wagging the model-theoretic tail. What our modal platonist needs is an accounting of models that is independent of antecedent logical principles. With such an accounting, we could compare the representative capacities of the already-existing and -known models with what we take to be the logical possibilities and determine whether the proposal is at least extensionally adequate. This, however, is exactly what we cannot do. We determine the nature of what models there are on the basis of accepted logical principles, both axioms and rules of inference. Our model-theoretic discoveries are driven by logical truths and consequences. Those with nonstandard ideas about logical principles will, understandably, have nonstandard ideas about the existence and representative capacities of the set of models. So, there is no independent appeal to models that can serve the typical anti-essentialist.

---

[10] Benacerraf, "Mathematical Truth," this JOURNAL, LXX, 19 (November 8, 1973): 661–79; Field, *Realism, Mathematics, and Modality* (New York Blackwell, 1989).

[11] For instance, C.C. Chang and H.J. Keisler in the classic *Model Theory* (New York: North-Holland, 1973), p. 2, state that "The three theorems we have stated [Löwenheim's, Compactness, and Morley's] also say something positive about the existence of models having certain properties. Indeed, in almost all of the deeper theorems in model theory the key to the proof is to construct the right kind of model."

To put this problem in another light, a methodological question is begged by modal platonism when it serves anti-essentialism. Why believe that the model-theoretic account of logical truth and consequence is correct? In particular, why should we believe that no model is impossible or represents an impossibility and why should we believe that the models cover all the possibilities? Well, because there are precisely enough of the right kinds of models, a modal platonist might say. Why should we believe this? The platonist might well go on to say that we should believe this because we discover the existence and nature of models by reasoning from fundamental axioms using fundamental logical principles; we construct or discover models in accord with the principles we take to be logically necessary. This gives the appearance, though, of solving a problem by doing nothing more than stating that there is a solution to the problem. When pressed about the form of that solution, that is, whether the class of models satisfies certain required modal constraints, the platonist simply says that the solution has an acceptable form, that is, yes, the class of models does meet constraints (i) and (ii). What is lacking is some access to the elements of the solution that is independent of the modality that is being investigated. The use of logical principles to infer the full character of the realm of models is unwarranted in the present context unless there is good reason to think that those principles do, in fact, track the possible accurately, which is precisely the point at issue between the essentialist and the anti-essentialist. Thus the anti-essentialist is not yet entitled to claim that the relevant biconditional is true.

The salient difference between the model-theoretic account of logical necessity and the discovery of legitimate reductions is that in the latter cases we can track the phenomena to be reduced independently of their respective reductive basis. For instance, we have known for millennia about the common phenomenon of combustion. What was not known was the nature of common combustion, that is, oxidation. As investigations continued, scientists gradually developed a theoretical and experimental framework within which we came to understand combustion as oxidation. Part of what legitimized the theory that common combustion is oxidation is that claims about the existence and nature of oxygen were incorporated into physical theory on the basis of much more than the need for an explanatory mechanism for combustion. If our only access to the process of oxidation were the mere fact that combustion occurs, then the explanation of combustion via oxidation would be an empty explanation. Thus, we can see that the model-theoretic account of logical necessity, in the context of the kind of anti-essentialism under discussion here, is empty.

The third problem is an ad hominem against the typical anti-essentialist who admits logical necessity. Part of the rationale for admitting logical necessity but not essentialism is that our knowledge of logical truth and necessity is supposed to be unproblematic, relying only on the semantics for a very few key (logical) concepts. Essences by contrast, are supposed to be real and objective, knowledge of which must depend on modal intuition. An anti-essentialist may well eschew a reliance on modal intuition because modal intuitions, like moral intuitions, are not universal and conflicts among intuitions are not easily resolved, if at all.

This is a false picture of the anti-essentialist's epistemic situation, however. If logical necessity is supposed to be a metaphysically significant necessity in the sense explained, that is, that the T axiom holds for logical necessity, then one adopting classical logic as encompassing all and only the fundamental necessary truths must face the fact that defending this commitment requires defending a substantial Humean metaphysical view. If the only necessities are austere logical truths expressed in primitive predicate notation, then whatever is not ruled out by those truths is possible; whatever is not contradictory is possibly true. This is simply a way of expressing in terms of grammatical form the metaphysical thesis that all logically possible combinations of basic properties are genuinely possible; there are no necessary connections between primitive properties. The problem here is not that the anti-essentialist will find this Humean metaphysics unattractive, but rather that the modal doctrine, the anti-essentialism, is hostage to a particular metaphysical program. This conflicts with the common perception that logic and its correlated modality are free of metaphysical commitment in a way that essentialism is not. If no question is begged against the essentialist, the anti-essentialist must rely on modal intuition in defending the preferred metaphysical framework no less than the essentialist. This reliance is unrecognized to the extent that we fail to acknowledge that our access to models is both dependent on our logical commitments and that these logical commitments must also be metaphysical commitments, if we take '$\Box_L$' to express a metaphysically significant modality. Of course, one may hold that '$\Box_L$' expresses a merely formal modality, but only at the cost of leaving it a completely open question what connection, if any, there is between logical possibility and real possibility. If one does not take logical commitments to be partly metaphysical commitments then, as I shall highlight a bit in section V, one will not have a straightforward explanation of the standard role of inference in both ordinary and philosophical contexts. The relative philosophical innocence of anti-essentialism vanishes.

IV. CONTEMPORARY CONSEQUENCE

The moral of sections II and III is that we cannot secure an adequate account of a metaphysically significant logical necessity if we limit ourselves only to the nonmodal facts of cognition, formal syntax, or model-theoretic structures. Either some underlying modality is presupposed or else there is no nonquestion-begging reason to assert the coextensiveness of the logically possible and the model-theoretic facts or the coextensiveness of the logically possible and the genuinely possible. There is no way to eliminate this, unless we find a very different epistemic access to models. Given the intimate connection between anything we might label "logical necessity" and the nature of logical truth and consequence, the moral is that anti-essentialist accounts of logical consequence fail. There may well be formal and epistemological features of the relation of logical consequence, but given that we take the validity of an inference to provide some kind of guarantee for the truth of the conclusion, given the truth of the premise(s), these features do not exhaust logical consequence. A modal component is crucial.

At this point our discussion connects with the current debate over the nature of logical consequence. In his critique of Alfred Tarski's account of logical consequence, John Etchemendy distinguishes representational and interpretational semantics.[12] In representational semantics, the models are representations of the different ways the world could be. Each line of the truth table, for instance, would provide a different way the world could turn out to be for the sentences under consideration. Each model represents a class of possible worlds. The interpretations of the sentences remain stable while the state of the world varies from row to row in the table. Interpretational semantics reverses this; it varies the interpretations of the sentences examined while leaving the state of the world constant. The various rows of the truth table tell us the semantic values of the relevant sentence when we interpret the atomic sentences in such ways as they come out true or false *in our world*.

According to Etchemendy, Tarski adopts the interpretational approach. Models are interpretations. Interpretations of what? Certainly not all vocabulary. If what we now take to be logical vocabulary admits of variable interpretation, then the account comes out wrong. The logical laws would have no stable interpretation and, so, would not be logical truths after all. Thus, we typically set aside some vocabulary to have stable interpretations: the logical constants. Truth according to all models, then, is truth according to all interpretations of the nonlogical vocabulary. What vocabulary merits constant interpretations? Tarski thought that the distinction between logical and nonlogical vocabulary was somewhat vague, but, "certainly not quite arbitrary."[13] Since he thought that there were "no objective grounds... which permit us to draw a sharp boundary between the two groups of terms" we should make the division so that the model-theoretic account of logical truth and consequence accords reasonably well with ordinary usage and be done with it. 'And' and 'all' qualify as logical; 'was president' does not.

If we grant that this procedure is satisfactory for classifying vocabulary (in)eligible for variable interpretations, we must still give satisfactory interpretations to the logical constants. Etchemendy's presentation of interpretational semantics limits interpretations to those on a fixed domain. Intuitively, whether there are finitely many objects in the domain is neither a logical truth nor a logical falsehood. Because

---

[12] *The Concept of Logical Consequence* (Cambridge: Harvard, 1990).

[13] "On the Concept of Logical Consequence," in *Logic, Semantics, Metamathematics*, J.H. Woodger, trans. (New York: Oxford, 1956), pp. 418–20.

etations on a given domain can introduce restric-
o problem interpreting sentences that are satisfied
ains. There is no way, however, for alternative interpre-
d the proper results for sentences that are satisfied only
omains. The domain of discourse serves to interpret the
s, which on any actualist account contains only actual ob-
uppose the domain is finite. Sentences formulated with no
gical expressions that are satisfied only on domains larger than the finite world would be, according to the interpretational approach, not just false but *logically* false, that is, true on no interpretation.[14] Even one who thinks that the actual domain is finite should be discomfited by this result. The size of the domain is contingent, so it is inevitable that any account of logical notions defined on such a contingency will be hostage to those contingencies and will be extensionally adequate only by accident, if at all. For this and related reasons involving the vagueness of the logical/nonlogical distinction, Etchemendy rejects the Tarskian account of logical truth and consequence.

Perhaps a more obvious and deeper reason for rejecting the interpretational approach here is that it cannot be the basis of a logical modality that permits contrafactual reasoning. Most of our reasoning about other possibilities concern not alternative interpretations of a language, but alternative ways the very same things we talk about with our language under its current interpretation might be. So, the only relevant models must be representational models.

Gila Sher and Stewart Shapiro each try to rehabilitate the model-theoretic account of logical consequence by rejecting Etchemendy's assumption that representational and interpretational semantics are mutually exclusive.[15] Each appears to adopt something like the modal platonism sketched above. Sher tries to preserve Tarski's claim that logical consequence is both necessary and formal and Shapiro gives an intuitive characterization of logical truth as truth in all possibilities under all interpretations of the nonlogical vocabulary, thus melding the representational and interpretational perspectives.[16] The reliance on both possibilities and interpretations incorporates features of both the representational and the interpretational approaches. The domain of interpretation, as well as the interpretative assignments to nonlogical expressions, is variable.

---

[14] Etchemendy, pp. 111–24.

[15] Sher, "Did Tarski Commit 'Tarski's Fallacy'?" *The Journal of Symbolic Logic*, LXI (1996): 653–86; Shapiro, "Logical Consequence. Models and Modality," in Matthias Schirn, ed., *The Philosophy of Mathematics Today* (New York: Oxford, 1998), pp. 131–56.

[16] Sher, pp. 667–83; Shapiro, p. 148.

At the same time they each wish their accounts to avoid, words, "such metaphysical nightmares as 'transworld identity crucial to note, however, that the modal element is an *intuitive* of their accounts of consequence. Both find its substance in the "adjustment" of two parameters: the designation of logical vocabulary and the strength of the set theory used in the meta-theory. According to Shapiro, "both 'world' and 'language' get varied, in an interrelated, controlled manner."[18] Since there is no issue of "finding" the interpretations of the typical logical constants in the models, their privileged status is secure.

The modality in the intuitive characterization of consequence is something that Sher and Shapiro take to be eliminable. "On my conception," says Sher, "formal possibility is reduced to mathematical existence and formal necessity to mathematical generality...."[19] Shapiro also makes it plain that he trades modality for ontology, endorsing Hilary Putnam's claim that "[m]athematics has...got rid of *possibility* by simply assuming that, up to isomorphism anyway, all possibilities are simultaneously *actual*—actual, that is, in the universe of 'sets'."[20] Worries about the cardinality of models are handled in both cases by incorporating an axiom of infinity.[21] Standard languages have the isomorphism property, that is, that a sentence is satisfied by a given model if and only if it is satisfied by all models isomorphic to the given model.

These moves certainly rescue the model-theoretic account of logical consequence from Etchemendy's criticisms of the interpretational version of that account. If the size of the domain is variable up to infinity, then it cannot be a necessary truth, much less a logical truth, that there are only finitely many objects. Neither Sher's nor Shapiro's account of logical consequence, however, is any use to the anti-essentialist. The various ways of interpreting their ideas were canvassed in

---

[17] Shapiro, p. 149; see also Sher, pp. 658–61.

[18] Shapiro, p. 149; Sher makes a similar point on pp. 675–76.

[19] Sher, p. 682.

[20] Quoted by Shapiro on p. 150, from Putnam, "What Is Mathematical Truth?" in *Mathematics, Matter, and Method: Philosophical Papers*, Volume 1 (New York: Cambridge, 1975), p. 70. In correspondence Sher has indicated that she accepts this position as well. It is worth noting here that Putnam's statement is ambiguous on whether possibility is eliminated in favor of truth-in-a-model or in favor of truth-according-to-a-model. Sher's and Shapiro's statements are likewise ambiguous, though the earlier quote from Sher appears to eliminate possibility in favor of truth-in-a-model.

[21] A relatively weak axiom is sufficient if only first-order logic is recognized as logic. A rather strong axiom is required if second-order logic counts as logic, since second-order languages permit the formulation of sentences satisfiable only in nonmeasurable domains.

the previous section and the deficiencies shown. If their trade of possibility for mathematical existence is a trade of modality for truth-in-a-model, then the modality is undefined over nonactuality. If they trade the logical modality for truth-according-to-a-model and refuse to impose conditions (i) and (ii) as constraints on what count as models and they take their proposals to be conceptual analyses, then the analyses fail. If (i) and (ii) are imposed, then the analyses are circular. If they eschew analysis and look only for a truth about the relation between logical truth and truth-according-to-models, then they are not entitled to maintain that they have found that truth. Model theory, then, offers no haven for safe modalizing.

V. WHAT ESSENTIALISM CAN DO FOR YOU

This is not the place for an extended defense of essentialism, but I will take this section to sketch how the essentialist can offer a genuine account of logical necessity where the class of logically necessary truths is a special case of truths about essences. What the anti-essentialist takes to be modally innocent semantic facts involving models, the essentialist sees to be closet essentialism about peculiar sorts of entities. Take again the simplest semantic structures: those represented by truth tables. We specify the meaning of the typical binary truth-functional connectives by delineating the four familiar rows in a truth table. In typical presentations of the semantics for truth-functional languages, this procedure is spelled out with no comment or else it is presented with the justification that the construction of truth tables is a combinatorial matter.[22] The construction of truth functional models is, however, not a strictly combinatorial procedure; it takes as its basis preferred fundamental logical principles. In classical logics that basis includes the Laws of Noncontradiction and Excluded Middle. Classical truth tables omit a row for the case in which the propositions are both true and false and they omit a row for the case in which they are neither. Like the more complex models of model theory, the logic drives the construction or discovery of models.

The presence of covert metaphysics—in this context, a theory about the possible—is evident when we, once again, consider contrafactual reasoning. Suppose that truth tables represent no modal information. Then their content is no more than that no proposition *is* both true and false and none manages to be neither. Truth tables that exhibit only this nonmodal information provide us with no reason whatsoever to think that had the world been different not all members of an inconsistent set of proposition would have been true. Reductio arguments would tell us nothing more about possibility than would a sound empirical investigation that resulted in the falsification of a simple empirical claim. Going no further than this leads us back to the problems of section III: '$\Box_L$' is undefined over contrafactual situations and truth tables provide no warrant for our usual inferences about contrafactual situations.

---

[22] Whenever the semantics for a propositional language is given, if it is provided without philosophical argument that in a truth table for formulae that are combinations of n atomic formulae the table must have 2^n nonidentical rows, this classical assumption is in operation. This recipe for truth-table construction is presented without comment in J.L. Bell and M. Machover, *A Course in Mathematical Logic* (New York: North-Holland, 1977), p. 21. The confusion of traditional assignments of truth values with combinatorial assignments is explicit in David Bostock, *Intermediate Logic* (New York: Oxford, 1997), p. 24.

Denying hidden modal import to standard semantic structures would involve rejecting as completely misguided some staples of the history of philosophy. The ontological and cosmological arguments for the existence of God depend upon modal claims, as do defenses of atheism which argue for the impossibility of the coexistence of God and evil, to take well-known examples. If facts of logic do not have implicit implications for the natures of either the bearers of truth value or the makers of truth value, then even if one successfully demonstrated that *God exists* and *Evil exists* constitute an inconsistent pair of propositions, in the available and very weak sense that not both of them are true, then no conclusions whatsoever about whether they are both true in other ways the world might have been would be thereby warranted. An atheist might be content to show only that *God exists* is false because *Evil exists* is true, but that was not the intent of the logical problem of evil. The strategy was to arrive at a conclusion about the actual world by going through a subsidiary conclusion about impossibility. No such strategy is available unless even elementary truth-functional semantics expresses modal information.

If the truths of logic as specified by truth tables are to be useable in our reasoning about contrafactual situations, then the semantic information contained in truth tables must represent not just the actual facts about propositions but also modal information about propositions (or some other favored truth bearers). This modal information, if it is to be useful in reasoning across the full range of the possibilities, must be information about the essence of propositions. Implicit in truth tables, then, is the thesis that what the tables represent are all and only the relevant possibilities for propositions. Conditions (i) and (ii) resurface as hidden assumptions in all elementary logical semantics.

Given that invoking semantic structures to explain logical truth requires that the semantic structures satisfy two modal constraints, what is the nature of the modality? We have seen that logical necessity as defined in terms of logical truths and consequences, which are in turn defined model-theoretically, can ill afford to have the requisite modality turn out to be logical necessity. Modalities typically defined in terms of logical consequence will be, for obvious reasons, even poorer candidates for the prior modality. Thus, the prior modality is most plausibly some kind of essentialism. If the semantics appropriate to a given logical apparatus reflects essentialism at least about propositions, directly or indirectly, then it is quite easy to see how, for instance, the logical problem of evil at least purports to yield a stronger result than an exhaustive empirical search into the existence of God and evil.

If logical necessity is a special case of essentialist necessity, then logical necessities form a subset of the metaphysical necessities. On the assumption that the only genuine possibilities are those acceptable to the essentialist, it follows that not all logical possibilities are genuine possibilities. What is needed now is some motivated rationale to account for the genuine possibility of only some of the logical possibilities. One such rationale is a completely general essentialist account of necessity, including logical necessity.[23] On such a theory, the fundamental notion is that of the essence of, the nature of, a thing. Essence is not further explained in terms of logical truth, logical form, or even possible worlds. The absolutely necessary truths, in the only sense of 'absolute' that essentialists care about, are those true in virtue of the essence or nature of the relevant items. Essentialists have at least two ways of explaining logical necessity. One is to explain logical necessities as those true in virtue of the natures of logical items, perhaps propositions and their constituent concepts or else other truth bearers and their constituents. Alternatively, logical necessities might be explained as those propositions true in virtue of the natures of every situation or every object and property, thus preserving the idea that logic is the most general science.[24] The nonlogical metaphysical necessities are those that lack this degree of generality. These necessities are those propositions true in virtue of the essences or natures of only some objects and their attributes. The commonly discussed essentialist theses purport to specify at least part of the essences of water, Kripke's lectern, and Queen Elizabeth that are not shared by other substances or objects. On this account, logical facts do, indeed, turn out to be general facts, not of syntactic, psychological, or semantic structures, but of essences.

---

[23] One version of which is presented by Fine in "Essence and Modality," and "Ontological Dependence."

[24] This is not to be confused with those propositions which are true *of* every situation or every object and its properties. A proposition may be true of an object, say, but not true in virtue of the nature of that object. Contingent truths are obvious examples, according to an essentialist. That universal truth is no guarantee of necessary truth has already been established on behalf of the essentialist.

Essentialist claims to the effect that something is logically possible but not metaphysically possible amount to the following: the features common to the nature of all things—familiarly expressed by logically consistent propositions—are insufficient to rule out a given possibility, for example, that Kripke's lectern might have been made of ice. When the full nature of that lectern is considered, however, this possibility is excluded. If the logically possible is simply that which contravenes no universal essentialist facts, then it is not at all surprising that not all logical possibilities are genuine possibilities. We cannot fully specify the way something is categorically if we limit ourselves to the categorical properties it shares with all other things. Likewise we cannot specify the nature of a thing if we limit ourselves to essential properties that it shares with all other things. Consequently, we cannot accurately specify the possibilities for something if we limit ourselves to the part of its nature that it shares with everything else. On the assumption that part of the essence of Kripke's lectern is not shared with everything else, its real possibilities are not a function of only that part of its essence that it shares with everything else. Because the essentialist claims that the lectern has more to its essence than what it shares with the essences of all other things, we will overgenerate possibilities for it if we ignore part of its essence, in particular the part about its chemical make-up.

This framework permits, though it does not demand, a quite general essentialist account of other metaphysically significant necessities. Physical necessities, for instance, might be seen as those necessities that pertain to all and only physical objects, thus making a complete science of the physical a complete specification of the essence of a physical object. Mathematical necessities are those that are true in virtue of the essences of mathematical objects. And so on.

What I have argued for here is only a general essentialist framework. I render no verdict on whether constitution is essential to water or Kripke's lectern and I make no commitment on whether genealogy is essential to Queen Elizabeth II. Nor do I render a verdict on whether the above approaches to physical and mathematical necessities are well advised. More specific versions of essentialism will rely upon more specific claims not directly relevant to the relation between logical necessity and essence. One line of consideration that will be relevant to the further development of essentialism is the matter of truthmaking and the role of individuals (as opposed to individual constants) and the role of properties (as opposed to predicates). To the extent that one gives a truthmaking role to individuals, properties, and/or constructs of such, quite general essentialism follows for them as well. For instance, one might think that linguistic facts are grounded in objectual facts.[25] In such a framework it is plausible to think that the propositions have their essences partly as a function of the facts they report and logical relations will be, partly, a function of essential relations among objects or states of affairs. Accordingly, propositions are the way they are essentially—at least with respect to logically relevant facts like consistency, logical consequence, and the like—because individuals, properties, and states of affairs are the way they are essentially. If this line is pursued, Noncontradiction and Excluded Middle come out as reflecting the essence of truth bearers (according to logical traditionalists) because these laws reflect the essences common to individuals and to properties. Those who deny the logical truth of Noncontradiction and/or Excluded Middle do so, ultimately, because they wish their logic to reflect their divergent views about what is and what is not essential to individuals and to properties and, hence, what is and what is not essential to the logically-relevant facts about truth bearers. Some differences in logic would turn out to be symptomatic of differences in metaphysics and not merely differences in meaning that float completely free of metaphysics.

VI. CONCLUSION

The privileged status that the commonly understood notion of logical necessity enjoys in much of analytic philosophy is unwarranted. It is under threat from the paraconsistentists of various stripes, on the one hand, who complain that it is unduly restrictive. It is under threat from the essentialists, on the other hand, who complain that it is unduly permissive as an account of a most basic possibility. Here I have argued that typical attempts to elevate logical necessity to a place of honor and security fail. Some other necessity must be taken as basic to provide for some appropriate constraints on the existence and nature of the class of objects that are thought to underwrite logical necessity, such as abstract models.

One can always bite one bullet or another: stick to actualist nominalism, for example, and hold that the possible does not outstrip the actual. This is a heavy price to pay for anti-essentialism. Perhaps eschewing all modality deserves a closer look, though this seems an exceedingly desperate alternative. What we cannot have is an attitude

---

[25] As does Sher, p. 670.

SCANLON ON PROMISSORY OBLIGATION: THE PROBLEM OF PROMISEES' RIGHTS*

When promises give rise to clear obligations, these can be accounted for on the basis of general moral principles.... —Thomas Scanlon[1]

Promises, along with their close cousins, agreements, are a ubiquitous feature of human life, and not surprisingly.[2] Promises and agreements are powerful tools. One aspect of their power is this.

Failing special circumstances, at least, if you have promised to see me at five, and nothing more has transpired between us, I have a special standing in relation to your actions. Should you remark that you are thinking of not showing up at five, I have the standing to insist that you do. Should you fail to show up, I have the standing to rebuke you on that account. By virtue of your promise, your showing up at five becomes my *right*.[3]

---

* Material related to the present essay was presented to the University of Connecticut philosophy department in the spring semester 2002, and to the University of Connecticut Humanities Institute in the fall semester of that year. I thank those present for their comments. I am grateful to Richard Arneson, Paul Bloomfield, David Brink, Christopher Clark, Scott Lehmann, David Slutsky, and John Troyer for discussions on other occasions. I dedicate this essay to the memory of Michael Robins (1941–2002).

[1] *What We Owe to Each Other* (Cambridge: Harvard, 1998), p. 315; hereafter WW.

[2] For some reason the philosophical literature has focused on promises rather than agreements. This may be because it has long been thought—wrongly in my view—that agreements are constructed out of promises On this see Margaret Gilbert, "Is an Agreement an Exchange of Promises?" this JOURNAL, XC, 12 (December 1993): 627–49, reprinted with some revisions in Gilbert, *Living Together: Rationality, Sociality, and Obligation* (Lanham, MD: Rowman and Littlefield, 1996), pp. 313–38. The precise nature of the relationship between promises and agreements is an interesting issue: my use of the phrase "close cousins" is not intended to beg any questions.

[3] This last point, which I take to be intuitive, is common ground between theorists of otherwise different persuasions with respect to the nature of rights. See, for instance, H.L.A. Hart, "Are There Any Natural Rights?" *Philosophical Review*, LXIV, 2 (April 1955): 175–91; Joseph Raz, "On the Nature of Rights," *Mind*, XCIII, 370 (April 1984) 194–214 (reprinted in Morton Winston, ed., *The Philosophy of Human Rights* (Belmont, CA: Wadsworth, 1989), pp. 44–60; page references here are to the original) Cf. Michael Robins, *Promising, Intending, and Moral Autonomy* (New York: Cambridge, 1984), p. 99. The qualification regarding special circumstances is intended to set aside, in particular, the cases of coerced promises or promises to perform immoral acts. I take it that these are best considered after one has an articulated understanding of less complicated cases In what follows, the reader should assume that coerced promises or promises with immoral content are not at issue. I discuss coerced agreements in Gilbert, "Agreement, Coercion, and Obligation," *Ethics*, CIII, 4 (July 1993): 679–706, reprinted in *Living Together*, pp. 281–312.

0022-362X/04/0102/83–109

As it happens, the philosophical literature has not focused on the way in which promising creates rights in promisees. Rather, the focus has been on how a promisor is obligated by a promise.[4] There is, evidently, nothing wrong with asking how promises obligate as, indeed, they seem to do. If you have promised to see me at five, and nothing more has transpired between us, you are surely in some sense obligated to do so. Something along these lines has been seen, indeed, as a self-evident or a priori truth.

Thus H.A. Prichard:

> Once call some act a promise, and all question whether there is an obligation to do it seems to have vanished.[5]

Intuitively, a promise obligates the promisor in a particularly direct way, and likewise directly gives rise to a right in the promisee. It is tempting to say that both the right and the obligation are part and parcel of promising.[6]

Since making a promise appears to involve little but the expression of the promisor's will, one might well think that promissory obligation is somehow directly willed into being. As David Hume observed, such an idea is "entirely conformable to our common way of thinking and expressing ourselves."[7]

How, then, do promises obligate? Thomas Scanlon's thoughtful treatment of this topic begins by discussing a popular type of account that he wishes to reject. Scanlon characterizes this type of account as a "two-stage" matter, the first stage being a "social practice" of promising (WW 295).

Here is a version of such a social practice account. First, there is a valuable social practice of promising. That is, there is a social norm or rule in a given society to the effect that one who does a certain thing under certain conditions (thereby "promising to do a certain thing") is to do something further (thereby "fulfilling the promise"). In addition, members of that society generally conform to this rule, and expect one another to do so, and the existence of the social practice as a whole promotes an important good. Second, in the circumstances comprising the first stage, it is morally wrong for members of the society in question to violate the relevant social norms.

---

[4] As can be seen from the titles of many articles—see, for instance, "The Obligation to Keep a Promise" (H.A. Prichard, circa 1940) and "Promises and Obligations" (Raz, 1982).

[5] "The Obligation to Keep a Promise," in Prichard, *Moral Obligation and Duty and Interest* (New York: Oxford, 1968), pp. 169–79, here p. 169

[6] If this is correct, it would seem that the special circumstances discussed in note 3 above either invalidate the promise—in other words, there is no promise in spite of appearances—or the promisor and promisee have the usual obligation and right, the special circumstances notwithstanding. I would argue for the latter alternative, but will not do so here—see Gilbert, "Agreements, Coercion, and Obligation," on the case of coercion.

[7] *A Treatise of Human Nature* (New York: Oxford, 2000), 3.2.5.3.

Suppose that these two stages have been reached in my society. All else being equal, it will then follow that if I say "I promise to feed your cat tomorrow," it would be morally wrong for me not to feed your cat tomorrow.

Scanlon does not directly attack "social practice" accounts, but argues in favor of an account of a different sort.[8] He proposes to argue that "the wrong of breaking a promise" is an instance of "a more general family of moral wrongs which are concerned...with what we owe to other people when we have led them to form expectations about our future conduct" (WW 296). He adds "Social practices of agreement-making, when they exist, may provide the means for creating such expectations...," but, he will argue, "these practices play no essential role in explaining why these actions are wrong" (WW 296).

I am inclined to agree with Scanlon that one should reject accounts of promissory obligation that appeal to social practices in the way described. One motive for developing such accounts has been the difficulty of understanding how—as our common way of thinking has it—the obligation of a promise can directly be *willed* into being. It is now standard to assume that another story about the genesis of promissory obligation must be found.[9] A problem with any other story, however, is that it is likely to be found counterintuitive.[10]

Its counterintuitiveness is not the only problem with an approach that appeals to social practices. Another problem is that such approaches are not clearly free, in their initial assumptions, of something akin to a "prior promise."[11] One needs to ask in what the existence of a social rule in a society consists: Might some kind of social acceptance of the rule be necessary, and might that involve something akin to a promise or agreement? If the assumption of anything akin to a promise is avoided, it may be harder to make the moral argument. If it is not avoided, then, at the least, more will need to be said about the thing-akin-to-a-promise that is involved in social rules.[12]

---

[8] Scanlon has developed this position in (at least) three places: "Promises and Practices," *Philosophy and Public Affairs*, XIX, 3 (Summer 1990): 199–226 (hereafter PP); "Promises and Contracts," in Peter Benson, ed., *The Theory of Contract Law* (New York: Cambridge, 2001), pp. 86–117 (hereafter PC), and chapter seven of *What We Owe to Each Other*. See also Scanlon's reply to John Deigh in Scanlon, "Reasons, Responsibility, and Reliance: Replies to Wallace, Dworkin, and Deigh," *Ethics*, CXII, 3 (April 2002): 507–28, pp. 522–28.

[9] One contemporary philosopher who has argued against this assumption is Robins—see, for instance, "The Primacy of Promising," *Mind*, LXXXV, 339 (July 1976): 321–40.

[10] This would hardly worry Hume, whose own account of promissory obligation in the *Treatise* appeals to a social practice along the lines indicated in the text. He famously argued that the idea of willing an obligation into being was unintelligible (3.2.5.3–7). His skepticism about the pretheoretical conception of promissory obligation has found favor among many Anglo-American philosophers. We should surely see it as both troubling, and challenging, however, if outside our studies, we understand the obligations of our promises to be willed into being.

It may be observed that, as sketched in the previous quotation, Scanlon's own account of promissory obligation itself has too many "stages" to accord with the intuition that promissory obligation is somehow directly willed into being. First, there is a certain set of circumstances, in part created by the promisor, that include, importantly, certain expectations of the promisee. Second, in the circumstances comprising the first stage, it is morally wrong for the promisor not to do what was promised.

Thus, though it does not invoke social practices, on Scanlon's account promissory obligations are not directly willed into being. This is one concern one might have with his account, as with any account sharing the feature noted.

I will focus here on a different concern: that Scanlon's account of promissory obligation does not explain the existence of promisees' rights. As I shall explain, if this is the case, his account of promissory obligation is inadequate.[13]

I. PROMISEES' RIGHTS: INTUITIVE POINTS

I.A. Hart on moral rights. In a classic article, "Are There Any Natural Rights?" H.L.A. Hart sets out an array of intuitive points on the nature of the rights of promisees. He refers to such rights as falling into a broad class of "moral rights." Precisely how he intended the qualifier "moral," need not detain us here. A central aspect of his intent is the differentiation of moral from *legal* rights. Certainly the rights of promisees are not in and of themselves legal rights, though coupled with special circumstances, such as the intent to create legal relations, they may give rise to legal rights.

---

[11] Cf. Prichard. The existence of this problem for a variety of theories of promissory obligation is a theme of Robins's "Primacy of Promising." For the relevance of the problem to Scanlon's theory, see the text, below.

[12] Observations in Hart, *The Concept of Law* (New York: Oxford, 1961), suggest that on a central everyday conception of a social rule something akin to a promise or agreement underlies social rules. I focus on these observations in developing a new account of social rules in chapter five of *Sociality and Responsibility: New Essays in Plural Subject Theory* (Lanham, MD: Rowman and Littlefield, 2000)

[13] Is a promisee necessary for there to be a promise? For present purposes, I waive this question. Like Scanlon, I focus specifically on the case of a promise with a (single) promisee, that is, someone to whom the promise is made, someone who can say "You promised *me* that you would. . ."

Hart lists the following features of moral rights in his sense: one *has* or *possesses* such a right, and corresponding to a right is another person's *obligation to* the right holder. This person *owes* the right holder what he (or she) has a right to: it is the right holder's *due*. If the obligated person fails to fulfill his obligation to the right holder he will have done *wrong* to the right holder. The right holder has a *claim* on the person who is obligated to him, and is *entitled* to have his right respected. He can *waive* the claim and *release* the person who is obligated to him from his obligation.

The emphasized terms—all emphasized in Hart's text—are central, closely related elements of what one might call the *language of rights*. Thus, Joe may *act wrongly* in relation to Jean. Or he may *do wrong to her, do her wrong*, or (even more briefly) *wrong* her. In the second set of cases, but not the first case, the words suggest that Joe has infringed a right of Jean's.

I.B. The "correlativity" of rights and (directed) obligations. The rights that Hart is talking about here are nonlegal analogues of what, in law, the eminent jurist Wesley Hohfeld called "claims." Hohfeld regarded claims as rights in the strict sense of the term.[14] In discussing Hohfeld, and consonantly with Hart, I shall refer to them simply as "rights."

Hohfeld insisted that there was a *duty* corresponding to every right. This duty was a duty *towards* the right holder, the person with the duty being the person the right was *against*.

There are, then, three elements that figure in the specification of each right and of the corresponding duty (in Hohfeld's terminology) or (in Hart's terminology) obligation. In the case of the right, there is a right holder, something the right is a right to, and someone the right is a right against. In the case of the duty, there is a duty holder, the thing it is a duty to do, and the person towards whom it is directed.

Duties or obligations of the type in question will be referred to as *directed* duties or obligations.[15] The duties or obligations corresponding to rights are, it may be said, intrinsically directed, as are rights themselves.

As is now common, I shall not here distinguish between duties and obligations.[16] The directed nature of certain duties (or obligations) will be indicated by the prepositions "toward" or more briefly "to." A general term that would seem to serve as well as the unqualified "duty" or "obligation" is "requirement," which I shall use on occasion. The "correspondence" between rights and directed duties is extremely tight. You cannot have a right without a corresponding directed duty, and vice versa. They are, if you like, two sides of the same coin.[17]

---

[14] "Some Fundamental Legal Conceptions," *Yale Law Journal*, xxiii (1913–14): 16–59, see especially pp. 30–32.

[15] This follows the terminology of deontic logic. Other qualifiers used include "relational."

I.C. Hart on special rights. In "Are There Any Natural Rights?" Hart distinguishes between *general* and *special* rights, the rights of promisees being "the most obvious cases" of the latter kind (*op. cit.*, p. 183). Special rights "arise out of special transactions between individuals or out of some special relationship in which they stand to each other" in such a way that "both the persons who have the right and those who have the corresponding obligation are limited to the parties to the special transaction or relationship" (*op. cit.*, p. 183). As to promising:

> We voluntarily incur obligations and create or confer rights on those to whom we promise; we alter the existing moral independence of the parties' freedom of choice in relation to some action and create a new moral relationship between them, so that it becomes morally legitimate for the person to whom the promise is given to determine how the promisor shall act. The promisee has a temporary authority or sovereignty in relation to some specific matter over the other's will which we express by saying that the promisor is under an obligation to the promisee to do what he has promised (*op. cit.*, pp. 183–84).

I.D. Hart on the character of directed obligations. The last sentence in the above quotation is of particular importance. As Hart sees it, being obligated towards another person is to be explicated in terms of that person's having a limited authority over one's will.

Hart sometimes writes of an obligation's being "owed" to the holder of the corresponding right.[18] Presumably what is owed is an action rather than an obligation. Thus Hart also speaks of a right holder as the person "to whom performance is owed or due."[19] He suggests, then, that an obligation of the type in question is a matter of the person with the obligation owing something to the right holder.

---

[16] Hart prefers to restrict the use of the term "obligation" for the directed duties that correspond to rights. He notes a broadening use of the term "obligation," however, such that "duty" and "obligation" are used interchangeably. See also Richard Brandt, "The Concepts of Obligation and Duty," *Mind*, LXXIII, 291 (July 1964). 374–95.

[17] For emphasis of this point and discussion of its implications for moral theory see Hugh Upton, "Right-Based Morality and Hohfeld's Relations," *Journal of Ethics*, IV, 3 (January 2000): 237–56. Roman law expresses the sense of such a tight relation by using one and the same word (*Obligatio*) for the duty of one party and the corresponding right of the other party to a contract. There was no special word for a right against a determinate person or persons. See George Long, "*Obligationes*," in William Smith, *A Dictionary of Greek and Roman Antiquities* (London. John Murray, 1875), pp. 817–21.

If I—somehow—owe you performance of a certain action, then there is a sense in which that action is already yours. In relation to that action, then, you are in a position to act towards me in a way appropriate to one who owns it.

This conception of things may help to throw light on a variety of aspects of the situation of a promisee. For instance, suppose a promisor asks his (or her) promisee for release from the promise. The promisee is in a position to grant this request, and thus to release the promisor from his obligation. If the promisee owns the promised action, then he is in a position to give it away, should he so choose.[20] Again, the promisee is in a position to keep the promisor's obligation in force. If the promisee owns the promised action, then he is in a position to keep it if he so chooses.

Pertinent here, also, is a statement of Hart's about rights in general prior to his discussion of promises. He says, "there is no incongruity, but a special congruity in the use of force or the threat of force to secure that what is someone's right to have done shall in fact be done; for it is in just these circumstances that coercion of another human being is legitimate" (*op. cit.*, p. 178).

Contemplating the exercise of "force" in connection with everyday promises and agreements may seem to go too far. It is reasonable, however, to include under the rubric of force the kinds of thing noted at the outset of this essay: such things as informal rebukes and demands. Informal rebukes lie at the thin end of the wedge of force: nonetheless, they are a form of authoritative chastisement.[21] Informal demands are a form of authoritative pressure to act in a certain way.

As applied to promises, one can take Hart's central point here to be about standing rather than justification. Whatever the merits of a forceful response in a given situation, if you fail to give me what I have a right to through your promise, I have the standing, as your promisee, to rebuke you on that account. Similarly, should you threaten to break your promise, I have the standing, as your promisee, to command or insist that you act as promised, and thus pressure you to perform.[22]

---

[18] See "Are There Any Natural Rights?" p. 179, note 7, also p. 181. Others sometimes write of "owing a duty."

[19] See, for instance, Hart, "Are There Any Natural Rights?" p. 181

[20] It may be that what the promisee has by way of ownership of the promised act is joint ownership with the promisor rather than sole personal ownership. He may then need the concurrence of the promisor in order to rescind the promise. Be that as it may, if the promisor has already sued the promisee for release, the latter is surely in a position "unilaterally" to release the promisor. I shall not go further here into the type of ownership that is in question.

[21] In his classic work, *The Concept of Law*, Hart associates such rebukes with the punishments imposed through the formal processes of law as versions of one form of human activity—see p. 10 and elsewhere.

Consider, now, a case of such insistence. Suppose that after lunch forgetful Fred casually remarks to his companion, Vera, "I'm off for a long hike!" and Vera authoritatively responds, "No, you're not!" explaining: "You promised me you'd help clean the house this afternoon!" If Fred's promise has brought it about that Vera in some sense owns his action of helping her clean the house, she would surely have the standing to insist that he does not withhold that action from her but rather performs it. She would also surely have the standing to rebuke Fred should he fail to give her what is hers.

In what follows I shall make the following assumptions about the directed obligations of promisors. In order to understand how a given promisor, *A*, comes to have an obligation *towards* his promisee, *B*, we need to understand how *B* comes to have a limited authority over *A*'s will, where this involves the promised action's being owed to *A*, its being already in a sense his. *B* will have the associated standing to effect the release of *A* from the promise if requested to do so or to hold him to it, and to command and rebuke *A* in the ways mentioned above. The same goes, of course, for our understanding of how *B* comes to have a right *against* *A* to performance of the promise. In what remains, I consider whether Scanlon's account of promissory obligation helps us to understand these things and hence, in effect, to understand the rights of promisees.

II. SCANLON ON PROMISSORY OBLIGATION

According to Scanlon, promissory obligations are a function of a moral principle that requires certain actions in specified circumstances. He labels this "Principle F" because he sees it as a principle of fidelity. He tells us that the conditions it specifies can be fulfilled in many ways other than by making a promise. "Promising is a special case, distinguished in part by the kind of reason that the promisee has for believing that the promisor will perform" (WW 306).

---

[22] Compare the description Scanlon gives of (claim-) rights in general as rights to "command" particular things, where others have a correlative duty to "comply" in "Rights, Goals, and Fairness," in Stuart Hampshire, ed., *Public and Private Morality*, (New York: Cambridge, 1978), pp. 93–111.

It appears, then, that in Scanlon's scheme a given promisor's obligation to do what he promised will exist, if it does exist, by virtue of the application of Principle F, as will the obligations of all who fulfill the relevant conditions. I shall not, therefore, spend time here on what, according to Scanlon, differentiates making a promise from acting in one of the other ways that fall under Principle F.

Let us assume that a given promisor will have an obligation—indeed, a moral obligation—that derives from Scanlon's Principle F. It is not at all obvious that this obligation corresponds to a right of the promisee against the promisor to performance of the promise.

In order for it to do so, it will have to be not just an obligation, but an obligation towards the promisee, an obligation that is the other side of the coin from the promisee's right against the promisor to performance. An obligation that exists by virtue of the application of a moral principle requiring certain actions in specified circumstances, though indeed an obligation, is not, or not necessarily, a directed obligation.[23]

One might sum up this line of thought by saying that there is an important challenge for any "moral principle" account of promissory obligation. Such accounts must show that what flows from the principle in question is not just an obligation, period, but a directed obligation.

At this point in the discussion someone may say "Whoa! What other kind of account can there be?" I believe that there is another kind of account, and a plausible one. In this article, however, I shall not go beyond the territory of moral principle accounts.

A variety of moral principle accounts is possible, and Scanlon's Principle F contains an interesting wrinkle such that it might be thought to be immune from criticism in terms of the rights of promisees. Scanlon may, indeed, have included this wrinkle in part so as to avoid criticism of this kind. There is some merit, therefore, in discussing the issue in relation to Scanlon's Principle F in particular.

I start with some observations on Scanlon's discussion preliminary to his presentation of his principle. In particular, I note some judgments on promises that help to motivate the principle.

II.A. Observations on promising. In the discussion preliminary to the introduction of Principle F, Scanlon introduces the "car and lawn" case:

---

[23] Concordant thoughts are found in Hart, "Natural Rights," pp. 180–81; also Joel Feinberg, "The Nature and Value of Rights," *The Journal of Value Inquiry*, IV (1970): 243–51, reprinted in Winston, ed., pp. 61–74, see p. 62

> Suppose, for example, that I promise to drive you to work if you will mow my lawn, and that you accept this arrangement. Then, a day or so later (but before the time has come for either of us to begin fulfilling the bargain) I think better of the deal and want to back out. On most people's understanding of promising, I am not free to do this. I am obligated to drive you to work unless you "release" me, even if I warn you before you have undertaken any action based on our arrangement. If I am going to break my promise then it is better to warn you than not to do so, but even if I do, this is a case of breaking a promise, not fulfilling one (WW 301, my emphasis; see also PP 205; PC 92).

Scanlon's emphasis here is on the fact that one cannot fulfill the obligation of a promise to do a certain thing by a timely warning to the promisee that one will not, after all, do that thing. He makes a similar point about compensation.

> If one fails to fulfill a promise, one should compensate the promisee if one can, but the obligation one undertakes when one makes a promise is an obligation to do the thing promised, not simply to do it or to compensate the promisee accordingly (WW 301).

In other words, having failed to do what one promised to do, one cannot bring it about that one has, nonetheless, fulfilled the obligation incurred by one's promise by compensating the promisee.

The quoted passages differ in relation to Scanlon's positive characterization of the obligation of the promisor. In the first passage, one is said to be obligated to do the thing promised unless "released" by the promisor. In the second, one's obligation is said to be an obligation to do the thing promised, period.

I take it that what is intuitively correct here is this. The content of the obligation of the promisor is, indeed, to do the thing promised, period. But that obligation stands as long as the promisee has not released the promisor from the promise. At this point, it seems reasonable to construe Scanlon's first positive characterization of the obligation of the promisor along these lines.[14]

II.B. Principle F. Scanlon concludes that in order to explain the obligations arising from promises: "we need a principle stating a duty specifically to fulfill the expectations one has created under certain conditions" (WW 302). The quoted words are somewhat ambiguous, but I take it that what Scanlon has in mind here is a principle stating a duty (or obligation) of the following form: if, under certain conditions, one has created expectations to the effect that one will do such-and-such, one must fulfill those expectations.

---

[14] Scanlon's statement that (in effect) a promisor is obligated "unless released" by his promisee may be thought to imply that action on the part of the promisee alone suffices to get rid of the obligation, whereas, according my (deliberate) characterization in this paragraph, action on the part of the promisee is necessary but may not be sufficient. For some congenial remarks on the insufficiency of action on the part of the promisee alone, see William Vitek, *Promising* (Philadelphia: Temple, 1993), pp. 101–02.

The principle Scanlon has in mind is understood to be a moral principle. When *is* a principle a *moral* one? The gist of Scanlon's own answer may be found in this statement of the core of his carefully argued and articulated form of "contractualist" moral theory: "an act is [morally] wrong if its performance under the circumstances would be disallowed by any set of *principles for the general regulation of behavior that no one could reasonably reject as a basis for informed, unforced general agreement*" (WW 153, my interpolation and emphasis).

I shall not attempt to discuss Scanlon's contractualism here. Nor shall I discuss his assumption that the moral principle that will explain promissory obligations has to do with *expectations* that have been induced in the promisee. Perhaps he does not so much assume this as infer it from the fact that he can, as he believes, explain promissory obligation in terms of such expectations.

After some further discussion Scanlon proposes "a principle of fidelity that requires performance" (WW 304), namely, Principle F:

> If (1) A voluntarily and intentionally leads B to expect that A will do X (unless B consents to A's not doing so); (2) A knows that B wants to be assured of this; (3) A acts with the aim of providing this assurance, and has good reason to believe that he or she has done so; (4) B knows that A has the beliefs and intentions just described; (5) A intends for B to know this, and knows that B does know it; and (6) B knows that A has this knowledge and intent; then, in the absence of special justification, A must do X unless B consents to X's not being done (WW 304).

At this point I am not going to question the truth of Principle F. Rather, I want to argue that, if it is accepted, it is not sufficient for an understanding of promisees' rights.

II.C. Scanlon's consent clause. Principle F is more nuanced than Scanlon's description of it as "a principle of fidelity that requires performance" suggests. It requires performance *unless the promisor consents to lack of performance.* I shall refer to the emphasized clause as the "consent clause."

How does the consent clause get into Principle F? In terms of the progress of Scanlon's discussion, it gets there as a reflection of the nature of the expectations that he takes to be at issue.

Prior to the presentation of Principle F, Scanlon writes that, in the car and lawn case, "the expectation you reasonably want to be able to form is the expectation that I will drive you to work *unless you consent to my not doing so*" (WW 302). The expectation Scanlon then takes to be at issue when one promises is the expectation that the promisor will do a certain thing unless the promisee consents otherwise.

Principle F reflects this characterization of the expectation at issue. According to the principle, when this expectation is produced in the appropriate conditions, the promisor must do the thing in question unless the promisee consents otherwise. The promisor will in this way act precisely in accordance with the promisee's—internally conditional—expectation.[25]

One suspects that Scanlon's consent clause—and, indeed, his characterization of the expectation in question—is also intended to reflect the fact that, as he wrote earlier of a particular promise, "I am obligated ...unless you 'release' me..." (WW 301). In other words, my "consenting" to your not doing the thing you led me to expect you to do, in the conditions specified by Principle F, is intended to be tantamount to my releasing you from your promise.

Scanlon puts the term "release" in quotation marks. I am not sure why, since the general idea of the promisee releasing the promisor from his promise is unexceptionable. I have just put the term "consenting" in quotation marks, though Scanlon does not do likewise. As I see it, the term "consent" may carry a load that is undesirable in this context. More specifically, it may seem to imply that I am in a position of authority in relation to you, something that would be true if you had, in effect, promised me something, or done something akin to that. This is presumably not something Scanlon either wants or needs to assume in the formulation of Principle F (where the term "consent" appears twice).

In what follows, therefore, I shall construe "consent" in such a way that my simply saying "I'm fine with your not doing it" or something to that effect is enough to count as my consenting to your not doing it. In other words, I can be in a position to consent—in this sense—to your not doing something even if I have no pre-existing authority over you with respect to your doing that thing.

III. WHY SCANLON'S PRINCIPLE F IS NOT SUFFICIENT FOR PROMISEES' RIGHTS

Let us suppose the conditions of the protasis of Principle F are fulfilled in relation to Anne (who is in *A*'s position) and Ben (who is in *B*'s position), and that "doing *X*" in this case is phoning Ben on Monday. Now suppose that Anne does not phone Ben on Monday, Ben did not consent to this, and there was no special justification for it. Does Ben have a special standing to rebuke her for not phoning him—a standing that is not generally held, or, at least, not held for the same reason? Again, were Anne to suggest to Ben, before Monday, that she did not plan to phone him, would he have a special standing to insist that she did—a standing that was not generally held, or, at least, not held for the same reason? If Ben had, in effect, a promisee's right to Anne's phoning him on Monday, then he would have both of these special standings.

---

[25] On the intended internal/external distinction as applied to intentions, see section VI below.

For simplicity's sake, I shall focus on the question of a special standing to issue a rebuke. I shall consider four possible grounds that Ben might be thought to have for rebuking Anne. For each of these I shall argue either that it does not, after all, give him the standing to rebuke Anne, or that it does not give him a special standing to do so. These grounds are: (a) Anne's violation of a moral principle; (b) Principle F's consent clause; (c) Ben's "right to rely" on Anne's performance; (d) Ben's being the intended beneficiary of Principle F.

III.A. Anne's violation of a moral principle. On what grounds might Ben rebuke Anne for not phoning him on Monday? The ground that Scanlon's principle most clearly suggests is this: *Anne has violated a moral principle: she has not done that which, according to Principle F, she must do in the circumstances.* What she had to do was phone Ben on Monday, unless he consented to her not doing so.

For present purposes there is a problem with a ground of this nature. Either it is liable to give the standing to rebuke Anne to many other people, as well as to Ben, or it gives nobody that standing.

What I have in mind is this. There appear to be the following two options as to who has the standing to complain just because a moral principle is violated.

First, there may be a general "right to do wrong" such that *no one* has the standing to rebuke a person merely on the grounds that he or she has *violated a moral principle.*[26] In that case neither Ben nor anyone else—on those grounds alone—has the standing to rebuke Anne for violating Principle F. It obviously follows that Ben has no *special* standing to rebuke Anne on these grounds.

---

[26] Jeremy Waldron has argued in favor of a right to do wrong. See Waldron, "A Right to Do Wrong," *Ethics*, XCII, 1 (October 1981): 21–39; see also William Galston, "On An Alleged Right to Do Wrong: A Response to Waldron," *Ethics*, XCIII, 2 (January 1983): 320–24. Waldron responds in "Galston on Rights," *Ethics*, XCIII, 2 (January 1983)· 325–27.

Second, there may be no general "right to do wrong" of the kind in question. In that case, it seems likely that *everyone* has the same standing to rebuke Anne merely on the grounds that she has violated a moral principle. Then, once again, Ben is in the same position as everyone else. He has no special standing in the matter.

Perhaps not absolutely everyone will have this equal standing in the second case. Perhaps some classes of people will be ruled out. It is implausible, however, to argue that *everyone but Ben* will be ruled out, if the basis for the rebuke that we are considering is simply the moral wrongness of what Anne has done.

It is an important question—both theoretically and practically—whether or not people generally have the standing to rebuke someone merely on the grounds that he has violated a moral principle.[27] There is no need to pursue that question here, however. For it seems to follow from the available positions that Ben has no *special* standing to rebuke Anne if we assume that his standing to rebuke her is grounded simply in the moral wrongness of what she has done, or, to put it differently, in her violation of a moral requirement.

This seems to be so however nuanced the requirement in question is. In particular, *the existence of Scanlon's consent clause appears to make no difference when matters are considered in this light.* That clause, embedded as it is within Principle F, only has the effect of qualifying what Anne must do, if she is to conform to the principle, or, in other words, if she is to act morally.

Does the consent clause really not give Ben a special standing to rebuke Anne? Perhaps its implications should be investigated more carefully.

III.B. Principle F's consent clause. Suppose Anne has failed to phone Ben on Monday, in the circumstances envisaged, and Ben later upbraids her for not calling him. Suppose, further, that it is common knowledge between Anne and Ben that each accepts Principle F, and each believes that no one has the standing to upbraid a person for simply contravening a moral requirement.[28] Anne asks Ben by what title he, in particular, sees fit to upbraid her in this case.

---

[27] I am myself inclined to the negative position. If one takes that position, and one believes, at the same time, that wrongdoers could do with rebukes from others, at times, one is likely to take a special interest in discovering how, in general, people gain the standing to rebuke one another that promises, among other things, seem to provide. This is a theme of Gilbert, "Shared Values, Social Unity, and Liberty" (manuscript).

[28] I use the phrase "common knowledge" in roughly the sense of David Lewis, *Convention: A Philosophical Study* (Cambridge: Harvard, 1969). For an elaboration of my own, see Gilbert, *On Social Facts* (Princeton: University Press, 1989).

Suppose Ben observes, with emphasis, "Well, it wasn't just anyone's consent that was at issue, it was *mine.* You were supposed to call me unless *I* consented to your not doing so, and I did not consent...." It seems that Anne can reasonably reply: "It is true that your consent was at issue in the way described. I don't see, however, how that gives you a special basis for upbraiding me for not calling you. Since you did not consent, I have, clearly, violated Principle F. In other words, what I did was morally wrong. We are agreed, however, that *that* does not give you the standing to upbraid me for what I did. I can't see how the fact that *your* consent was at issue does."

Anne might go on: "Consider the following case. Jane promises Diana that she will stay in the house with Timmy that night unless he consents to her not doing so. Later that night she leaves the house without Timmy's indicating that this is fine with him. She has broken her promise to Diana, and, intuitively speaking, Diana, the promisee, has a special standing to rebuke her for so doing. What about Timmy? Does he also have a special standing to rebuke Jane? I don't see that he does. Where would this standing to rebuke come from? Not, as far as my story goes, from her breaking a promise to him, or doing anything analogous. After the fact, he can certainly observe, as you, Ben, can observe: 'It was *my* consent that was at issue, and I did not consent....' But I do not see how that gives him a special standing to rebuke Jane, or to demand that she stay home when she prepares to go out."

Suppose Ben responds: "Consider this, then. Timmy had the power to bring it about that Jane lacked an obligation she otherwise had—the obligation to stay in the house with him on the night in question. All he had to do was consent to her not staying, and she would no longer have an obligation to stay. Similarly, by withholding his consent he was able to keep her obligation in force. In short, he had the power to release her from her obligation as a promisor—or hold her to it. Such power is a function of the authority that gives a promisee the standing to rebuke a promisor for failing to fulfill his promissory obligation. So Timmy has that standing—and so, by parity of reasoning, do I."

Anne could surely reply: "Jane's obligation through her promise is conditional. Timmy cannot release her from that conditional obligation, or hold her to it. Diana can. In other words, if Jane pleadingly asks Diana, '*Must* I stay with Timmy unless he consents?' Diana's saying, 'No, you need not,' brings Jane's conditional obligation to an end. Timmy does not have the power to do that. Timmy's power is only the power to fulfill or fail to fulfill a condition of Jane's conditional obligation. That he has this power does not imply he has the standing to rebuke Jane if she fails to fulfill her obligation."

III.C. Ben's "right to rely" on performance. Scanlon says that "the obligation to keep a promise is *owed to* a specific individual" (WW 316; emphasis mine). This may be thought to imply that the individual in question has a special right to performance of the promise, and, consequently a special standing to rebuke the promisor. Scanlon may or may not mean his reference to what is *owed* to imply this.[29] However, he does seem to allude to such a special right when he says the following:

> When the conditions of Principle F are fulfilled, it would be wrong, in the absence of special justification, for the party in A's position not to perform. In addition, the party in B's position has a "right to rely" on this performance: that is to say, the second party has grounds for insisting that the first party fulfill the expectation he or she has created (WW 305).

I am not sure why Scanlon puts quotation marks round the phrase "right to rely." In any case, he evidently—by his use of the words "that is to say"—wants us to interpret this phrase in terms of grounds for insisting that the first party fulfill the expectation he or she has created. In other words, he interprets a "right to rely" on a performance as, in effect, a right to that performance itself.

There is another way of interpreting the phrase "right to rely," a way such that having a right to rely is not equivalent to having a right to the thing relied on, and does not, or at least not obviously, entail having such a right. According to this interpretation, "having a right to rely on a performance" means the same as "being epistemically justified in thinking that a performance will occur (and hence to act as if it will occur)."

I am not sure that fulfillment of the conditions of Principle F entails that—in the case of Anne and Ben—Ben *will* be epistemically justified in thinking that Anne will perform as expected unless he consents to her not so performing. For present purposes I shall assume that it does. While there is, in that case, no doubt that Ben has a right to rely on Anne's performance in the epistemic justification sense now in question, one must distinguish between this kind of right to rely and a right to rely that is equivalent to, or entails, a right to the performance itself, a right that gives one a ground for insisting on performance. I doubt that Scanlon was under any other impression, but it is worth pointing out the possibility of making an inadvertent slide from a claim that a right to rely in the first sense is present to a claim that a right to rely in the second sense is present.

---

[29] For some observations that suggest that he does not, see the discussion of Scanlon's use of "owing" in Frances Kamm, "Owing, Justifying, Rejecting," *Mind*, cxi, 442 (April 2002): 323–54, at pp. 333–36.

Is it worth pursuing, in the present context, the idea that one might somehow argue from a right to rely in the epistemic justification sense to a right to rely in the right to performance sense? I think not, in part for this reason. In the present case, an observer of the situation may, presumably, have as good an epistemic justification as Ben for expecting Anne's performance. If an observer—Ben's trusty bodyguard, for instance—can for that reason have a right to Anne's performance, this will not be an argument to the effect that Ben has a special right to Anne's performance, a right he holds by virtue of being the person to whom something analogous to a promise was made.

III.D. Ben as beneficiary of Principle F. Suppose Ben says: "I am—surely—the intended beneficiary of the duty specified in Principle F, as that principle applies to you and me. Hence I had a right against you to your fulfillment of this duty. This right gave me the standing to rebuke you for not fulfilling the duty, a standing that, in this situation, is special to me."

Anne might surely reply: "I will agree, for the sake of argument, that there is a sense in which you are the intended beneficiary of the duty specified in Principle F. I would still draw the line at saying you had a right against me to my fulfillment of the duty. To my mind, it is possible—without contradiction—to accept that I have acted contrary to a moral requirement that was in some sense intended to benefit you, while not accepting that you have a right against me to my acting in accordance with the requirement in question."

Ben appealed, in effect, to a *beneficiary theory* of rights. According to his version, the intended beneficiary of someone's moral duty is the holder of a right against that person to the fulfillment of the duty. Related theories of rights are quite popular, so it would be well to go into this matter further here.[30] In the next section, I briefly consider one such theory in relation to the question of a promisee's rights.

---

[30] Often "beneficiary theories" are contrasted with "choice theories" and both seen as theories of the function of rights. However, in some cases (see the text below) the notion of a beneficiary is used to explain what a right is, and how to determine whom a given right is a right against.

IV. BENEFICIARY THEORIES AND PROMISEES' RIGHTS

A distinguished exponent of a type of beneficiary view of rights is Joseph Raz.[31] In a well-known discussion, Raz prefaces his account as follows:

> ...a philosophical definition of "a right" like those of many other terms, is not an explanation of the ordinary meaning of a term. It follows rather the usage of writers on law, politics, and morality, who typically use the term to refer to a sub-class of all the cases to which it can be applied with linguistic propriety (*op. cit.*, p. 194).

Raz may mean to suggest, plausibly, that terms can get used so widely in everyday discourse that philosophers and others should be comfortable developing partial accounts of their referents, accounts that pick out an important sub-class of everyday uses. And one can, of course, introduce any term or phrase, including "a right," in any technical sense one pleases. This is something that philosophers, among others, often do. If, however, one is specifically interested in the rights of promisees, one would presumably want to give an account that was consonant with an intuitive characterization such as that given by Hart.

Raz presents the following fairly complex definition:

> "x has a right" if and only if x can have rights, and, other things being equal, an aspect of x's well being (his interest) is a sufficient reason for holding some other person(s) to be under a duty (*op. cit.*, p. 195).

This may capture a prevailing use of the phrase "a right" among writers in the areas Raz specified, as he intended. It has certainly had some influence over the way in which rights are understood in moral theory.[32] However, it is hard to see how it captures the nature of the rights of promisees.

A central thought about a promisee's right that seems to have no real place here is the thought that the person(s) who are under the duty referred to *owe* the action that fulfills it to the so-called right holder. There seems to be a logical gap between its being the case that I have a (moral) duty whose ground is an aspect of the right holder's interest, and its being the case that I owe performance of that duty to the right holder. It seems that it would be perfectly intelligible for me to say: "Okay, this duty I have is grounded in an aspect of your interest...but I do not *owe* you its performance. It's just my duty, period."

---

[31] See "On the Nature of Rights." I here include under the broad label "beneficiary theory" all those theories that make central to their account of rights the benefits, intended benefits, or interests of the putative right holder. Such theories may differ in important ways among themselves. Raz notes that his definition of a right "draws on several elements of analyses of rights which stem from Bentham's beneficiary theory" (p. 195, note 3). Among contemporary theorists, Raz cites Ronald Dworkin, *Taking Rights Seriously* (London: Duckworth, 1977), p. 100; D.N. MacCormick, "Rights in Legislation," in P.M.S. Hacker and Raz, eds., *Law, Morality, and Society* (New York: Oxford, 1977), pp. 189–209; and, especially, Keith Campbell, "The Concept of Rights" (D.Phil. thesis, Oxford University, Bodleian Library, 1979).

[32] See, for instance, Richard Arneson, "Against Rights," *Philosophical Issues*, XI (2001): 172–201, see third paragraph.

If I were to allow that I owe you performance, then I would recognize that you have the standing to upbraid me for nonperformance, or to insist on performance. Before the fact you could pressure me, saying in effect: "Give me that! It's *mine!*"

It is true that a bystander could say, "Give her that! It's *hers!*" The case in which I command you to give me what is *mine* is special, however. The bystander's standing to command you to give it to me can be questioned. My standing surely cannot be questioned. In the bystander's case, the riposte "It's none of your business!" makes sense. In my case, it does not.

In a discussion of promises that follows his definition of "a right," Raz focuses briefly on the rights of promisees. He seeks an account of these according to which there is such a right corresponding to every promise. He therefore favors the view according to which, though anyone may lose interest in having a given promisor fulfill his promise, each person has a general interest that promises made to him shall be kept. This is the interest "to have voluntary special bonds with other people" (*op. cit.*, p. 203).

Of the promisee he says that "It is always up to him to waive his right under the promise and thus terminate the binding force of the promise" (*op. cit.*, p. 203), in other words, terminating the duty of the promisor. It seems that he might have suggested, though he does not, that the promisor's duty to perform is subject to a consent condition akin to Scanlon's.

There is nothing in Raz's discussion so far that seems to imply that a right is held *against a certain person* or that it corresponds to a *directed* duty. Nonetheless Raz says that "Rights are held against certain persons" (*op. cit.*, pp. 209–10). In his eyes, this seems to come down to the question of who can satisfy the interest that grounds the right. Thus, "since contractual rights are based on an interest in being able to create special relations, they give rise to rights against other parties to the agreement as they are the only ones who can satisfy that interest on that occasion" (*op. cit.*, p. 210).

If one wishes, one can stipulatively define "right R is held *against person P*" in some such way. Nonetheless, the points already made suggest that where rights are held against persons, in *this* sense, the right holder does not necessarily have the special standing associated with the rights of promisees as ordinarily understood. I conclude that the idea of explicating rights of the latter kind in terms similar to those of Raz is problematic, and shall not pursue it further.

V. ADDITIONAL PRINCIPLES

Suppose it is accepted that Scanlon's Principle F does not in and of itself account for the rights of promisees. It may be suggested that Scanlon can plausibly account for these rights by introducing some kind of subsidiary principle, in addition to Principle F. I think this is doubtful, as I shall argue with respect to two rather different kinds of proposals that might be made in this regard.[23] I consider these proposals in relation to the example of Anne and Ben.

V.A. A social rule in the context of Principle F. The first kind of proposal invokes a social rule. A version of it might run roughly as follows. Assume that situations in which Principle F applies are common within a particular society, *S*. It is predictable that *S* will adopt a rule for behavior in such situations—call it *Rule R*—which has the following implication in the case of Anne and Ben: if Ben chooses to speak in a rebuking tone to Anne for not phoning him on Monday, she is not to object to his doing so on the grounds that he lacks the standing to rebuke her. In other words, she is to behave as if he has such standing. Indeed, she is to behave for all the world as if he has a right to her phoning him. The existence of Rule R in *S* would tend to promote the interests of those members of *S* who find themselves in Ben's position, interests that Principle F itself promotes. The ubiquity in human societies of Rule R would help to explain the intuitive judgment that promisees have the standing to rebuke promisors, and, indeed, the intuitive judgment that this standing derives from a right against the promisor, who owes the promised performance to the promisee.

To this proposal the following objection may be raised. Whether or not it is apt in some sense to explain them, the ubiquity of Rule R would not seem to *justify* the intuitive judgments in question. Rule R requires Anne to act *as if* Ben has a certain right. Intuitively, however, he actually has such a right. If what we are looking for, then, is an account of our intuitive judgments that provides a justification of them rather than a debunking explanation, this proposal must be rejected.

One who accepts this aim may counter that the existence of a rule in one's society to the effect that in certain circumstances one is to be treated as if one has a certain right is a sufficient basis for allowing that, in that society, one actually has that right. Or a rule understood in somewhat different terms may be invoked: if, in a given society, it is allowed that a promisee is to *count as* having a right to performance against the promisor, then, in that society, the promisee has such a right.[34]

---

[23] In informal discussion of an earlier version of this paper, the first was suggested by Arneson, the second by David Brink (La Jolla, November 2002).

There is a familiar problem common to both of these responses: they are not clearly free of the assumption of a "prior promise." What, after all, does the existence of a social rule in a given society amount to? According to at least some accounts, something akin to a prior promise or agreement is among the existence conditions of a social rule.[35] Thus these responses are open to the objection that they are attempting to explain a puzzling feature of promising by reference to something with that very feature. Suppose we set this problem aside, and allow, at the same time, that the existence of an appropriate social rule would actually give Ben a right against Anne. This still leaves the proposal at issue with serious problems. In particular, it detaches promisees' rights from promises, and from promissory obligation, in an unintuitive way.

Intuitively, a promisee's rights are part and parcel of promises. They are on a par, in this way, with a promisor's obligations. According to this proposal, however, a promisor's obligation may exist when a promisee's right does not: that right depends on the existence in the promisee's society of an appropriate rule. To say this is to say something in the spirit of Scanlon's doubts about practice theories of promissory obligation: promises and the like appear to obligate independently of any background social practices, rules, or understandings as to what counts as what. In the same way, they give promisees rights.

Relatedly, from an intuitive point of view, a promisee's right is the obverse of a promisor's obligation (and vice versa), a promisor's obligation being, by its nature, a *directed* obligation. The proposal at hand, meanwhile, starts with a nondirected obligation—that which is a result of Principle F—and puts it together with a social rule that is supposed independently to generate a right. Clearly this right is not simply the obverse of the obligation in question.

The proposal under consideration, then, has several problems. These are the problem of whether and how a social rule could give a promisee a genuine right, the prior promise problem, and the structural problem. Even if the first two problems are set aside, the relationship between an obligation derived from Principle F and a right supposedly the product of a social rule is not structurally the same as the intuitive relationship between a promisee's right and a promisor's obligation.

---

[34] The formula "*X* counts as *Y*" plays a large role in the account of social institutions in John R. Searle, *The Construction of Social Reality* (NY. Free Press, 1995). See also Searle, *Speech Acts: An Essay in the Philosophy of Language* (New York: Cambridge, 1969).

[35] See Gilbert, *Sociality and Responsibility*, chapter 5.

V.B. Subsidiary moral principles. Rather than appealing to a contingent social practice, can Scanlon plausibly appeal to a subsidiary moral principle? A beneficiary theorist of rights may point out, as did the proponent of the previous proposal, that it is in the interests of someone in Ben's position to be able with impunity to act as if he had a right to Anne's performance. Might one not argue, indeed, that Ben's interest in being able so to act is sufficient to ground a moral duty in Anne, and that Anne has, therefore, a moral duty to allow him so to act?

What exactly is Anne's duty according to this argument? It seems to be a duty to allow Ben to act as if he had a right to her performance. Does this mean that Ben does in fact have a right to her performance? That is at best not clear.

Perhaps someone will say that if certain others morally ought to act as if I have a given right against them, then, whatever else is true, I have a *moral right* against them with the same content. Perhaps it will be said that that is *what it is* to have a moral right—or one type of moral right—against certain others.

Suppose that for the sake of argument one accepts these points. One must then observe that a moral right of the kind in question is, intuitively, of a different type to the rights of promisees. Promisees' rights are not, at base, only rights-people-morally-ought-to-act-as-if-the promisee-had. They are—in addition—rights the promisee has. The same can be said of similar understandings of what a moral right is. It can be said, for instance, of the understanding that to have a moral right against someone is for it to be the case one morally ought to have a right against him. A promisee's right is a right the promisee has.

Another problem with this proposal has to do with the relationship between a promisee's right and a promisor's obligation. Once again, we have a proposal that starts with a nondirected obligation—unlike a promisor's obligation—and adds to the mix a type of right of the promisee—a right-people-morally-ought-to-act-as-if-the-promisee-had. Though it may not be correct to say that the right of the promisee is in this case only contingently present when the obligation of the promisor is, it is not simply the obverse of the promisor's obligation. It is something additional. For the promisor's obligation derives from Principle F, whereas the promisee's right relates to a different, here unnamed, principle, concerning an interest of Ben's that—according to the principle—morally ought to be protected.

VI. PRINCIPLE F AS THE SOURCE OF A SECONDARY OBLIGATION PRESUPPOSING PROMISEES' RIGHTS

There could be more than one obligation, and perhaps more than one *type* of obligation, associated with promises. I take it that, first, there is a (primary) obligation that is part and parcel of promising, a directed obligation correlated with a right of the promisee against the promisor. Then there could be one or more other obligations, always or perhaps only sometimes present when promising occurs. These (secondary) obligations would perhaps reflect the existence of the obligations that are part and parcel of promising. The secondary obligations would not necessarily be directed obligations with correlative rights.

The gist of my argument so far has been that Scanlon's Principle F is not the source of the primary obligation, given the directed nature of that obligation or, in other words, its correspondence to a right in the promisee. Principle F could, however, be the source of a secondary, nondirected obligation that obtains in some or all cases of promising.

There is, indeed, reason to think that Principle F is least debatable, as a moral principle, if one of its assumptions is the making of what is, in effect, a promise, complete with the primary directed obligation of the promisor. This may be argued roughly as follows.

Scanlon's statement of the principle starts with the assumption that a person *A* voluntarily and intentionally leads another person *B* to expect that *A* will do *X* (unless *B* consents to *A*'s not doing so). But there are many ways in which *A* might lead *B* to believe this. Thus *A* might express an appropriate personal intention. But *A*'s personal intention is surely *A*'s to change as *A* wishes. If all that *A* has done in bringing about the truth of the assumption is express a personal intention, then it is not at all clear that there will be any moral onus upon *A* to the effect that *A* must do *X* unless *B* consents to *A*'s not doing so.

Scanlon's second assumption is that *B* wants to be "assured" that *A* will do *X* (unless *B* consents to *A*'s not doing so), and his third is that *A* acts with the aim of providing this assurance, and has good reason to believe that this has been done. Now it may be that a simple expression of personal intention is not enough to *assure* anyone of one's future actions, because one may always choose to change one's mind. In that case, it seems that *A* must have done more than express such an intention.

In Scanlon's discussion, he at one point refers to a "settled intention," but it is not clear on the face of it what a "settled" intention is, other than a genuine one. However genuine my intention, I am surely at liberty to change my mind when I wish to, if my intention is the only relevant constraining factor. So the expression of a genuine intention is unlikely to be enough to assure someone that I will indeed do the thing intended. Rather, I will do it unless I change my mind.[36]

What is the appropriate intention here? Presumably it will incorporate a consent clause. Might the inclusion of this clause make a difference, constraining *A* from changing course?

Consider the statement "I intend to do *X* (unless you consent to my not doing so)." The stated condition, your consenting to my not doing so, may in principle be construed as internal or external to the intention. What I have in mind is this: when it is construed as internal, the quoted sentence comes out as equivalent to "I intend this: to do *X* (unless you consent to my not doing so)." When construed as external, it comes out as "Unless you consent to my not doing *X*, my intention is to do *X*."

As to the first construal, there seems no bar from the consent clause to my changing my mind, and canceling or otherwise amending my intention. To cite two possibilities, the intention that results from my change of mind might be the intention not to do *X* whether or not you consent to my not doing so, or to do *X* whatever you say.

As to the second construal, it seems to imply that I have an intention to do *X* that will persist unless you consent to my not doing *X*. It is not clear, however, how a personal intention can become subject to another person's consent. Is my claim that my intention will persist absent your consent an empirical one? Then how can I be sure it is true? Is it or does it presuppose an undertaking of some kind? If so, then it seems that there is, in effect, a promise in the background, a promise complete with a directed obligation of the promisor and its correlative right in the promisee.

In short, it is not easy to see how the assumptions with which

---

[36] A similar drift, not directed at Scanlon in particular, is to be found in various writings—see, for instance, Prichard, "The Obligation to Keep a Promise"; Robins, "The Primacy of Promising"; and, in relation to Scanlon, Michael Pratt, "Scanlon on Promising," *Canadian Journal of Law and Jurisprudence*, XIV, 1 (January 2001): 143–54, of which I learned after formulating similar points (Gilbert, "Moral Obligation and Agreement" (manuscript), presented at the University of Connecticut/Storrs, spring semester 2001). A focus of previous writing of mine has been the unilateral rescindability of personal intentions and other sources of personal commitments. See, for instance, my 1997 essay, "What Is It For *Us* to Intend?" reprinted in Gilbert, *Sociality and Responsibility*, pp 14–36.

Scanlon introduces Principle F lead to a *non*directed obligation of the person in *A*'s position unless they implicitly include what is, in effect, a promise of *A*'s to do *X*, complete with the directed obligation of the promisor. If we assume—contrary to Scanlon's drift—that this is so, then Principle F may hold. I shall not attempt to decide if it does here. If it holds when and only when these amplified conditions are met, then it will indeed describe an obligation that promisors often, if not always, have. It will not, however, describe the primary, directed obligation of a promisor.

VII. SUMMARY, CONCLUSIONS, PROSPECT

Scanlon's account of promissory obligation is one of a certain general type, a type that appeals to a moral principle applied in a noninstitutional context. It is intended to apply on the basis of facts about personal intentions, expressions of such intentions, expectations, and preferences. I have briefly argued, concordantly with other authors, that in order to be valid such a principle needs a different basis. It may well need a promise or something like it as part of the basis on which it will apply. Be that as it may, my central argument in this essay has been that Scanlon's principle is not equipped, in and of itself, to account for promisees' rights and the correlative directed obligations of promisors. My argument strongly suggests, if it does not show, that no account of this type will adequately account for the nature of a promisee's rights as these are intuitively understood.

If so, and if one agrees with Scanlon that social practice accounts are implausible, one might wonder: How are a promisees' rights possible? And if it is hard to see how rights of this nature are possible, their actuality may be doubted. One might then be tempted to stick with a noninstitutional, moral principle account of promissory obligation such as Scanlon's. And one may give up the search for a source of a promisees' rights.

I think this would be a mistake, since I believe that it is possible to give an account of promisees' rights that is neither a practice account nor a noninstitutional, moral principle account. I shall not describe nor discuss the account I have in mind here.[37] I shall, however, make some points relevant to it.

---

[37] I have sketched such an account in various publications. These focus on agreements rather than promises, but suggest an interpretation of promises as well. See, for instance, my articles, "Is an Agreement an Exchange of Promises?" "Agreements, Coercion," and "Obligation and Joint Commitment," *Utilitas*, xi, 2 (January 1999): 143–63, reprinted in Gilbert, *Sociality and Responsibility*, pp. 50–70. I return to it in my book, *Rights Reconsidered* (to be published by Oxford University Press). Robins, in *Promising, Intending, and Moral Autonomy*, emphasizes the need to account for a promisee's rights (at pp. 99–102) and pays careful attention to that need in developing a proposal as to the nature of promising (see pp. 102–03). My own proposal has some affinities

On the topic of promissory obligation, people may seesaw between noninstitutional moral principle views and practice views that, also, appeal to a moral principle because they take seriously Hume's worry that there is no sense in the idea of willing an obligation into being. Now this may seem to be absolutely clear if we are thinking of an obligation, duty, or requirement that derives from a moral principle, a principle, roughly, of the type that Scanlon's contractualism attempts to characterize. Call such obligations, and these alone, *moral* obligations. One may think, plausibly: How could anyone will a moral obligation into being, just like that, just because they wanted to? A moral obligation simply is *not* that kind of thing. One has or does not have such an obligation, in the circumstances. Given such-and-such circumstances, the obligation is or is not there. In a given context one may be able to create the relevant circumstances by an exercise of will, but that is not willing the *obligation*.

Supposing this is right, it need not end discussion of the possibility that the directed obligations of promisors are directly willed into existence. For it could be that the obligations associated with promises, while being genuine and indeed paradigmatic obligations, are not moral obligations (nor legal obligations either).

Though one might be unclear as to how this could be, it can surely not be thought unintelligible. That there are nonmoral, nonlegal obligations and rights would only be unintelligible if by "moral" one meant "other than legal." Then the categories of "moral" and "legal" would be exhaustive, by definition. But a moral obligation is not being conceived of here in such a residual fashion, but rather in terms of an intuitive idea of the moral realm. Perhaps a sense of the problematic nature of moral principle views, along with practice views, will allow those otherwise skeptical of a *via media* to view such a possibility more favorably.

Evidently, the discussion here is relevant to more than the rights of promisees and the correlative obligations of promisors. It bears on the general question of the nature and source of rights in the sense of Hohfeldian claims or rights "in the strict sense of the term." When their directionality is construed as in this article, such rights—for short, *rights*, period—are clearly valuable possessions. If I have a right against you, I have the standing to put pressure upon you to do certain things. I have a certain, if limited, authority over your will. How, then, do people come by such rights?

---

with that of Robins—for whom the notion of *commitment* is key. There are important differences too, in particular, my invocation of a commitment that is *joint* in the sense I articulate.

Insofar as legal rights are *stipulated* rights, they raise immediate questions about the authority of law itself. The same goes for any other stipulated rights, including the rights, if any, that are created by nonlegal rules. Setting such rights aside, many moral philosophers have been inclined to suppose that some rights—which they would refer to as *moral* rights—are constructed out of purely moral materials such as moral requirements. These philosophers do not think of moral rights as rights that ought to be, or some such thing. They think of them as real, existing rights constructed out of purely moral materials. The discussion in this article suggests that there are no moral rights in this sense.[38]

If that is right, it leaves us with the question: Where do nonlegal, nonstipulated rights come from? One source of such rights is, evidently, the source of promisees' rights. Clearly, it would be good to know what that is.

MARGARET GILBERT

University of Connecticut/Storrs

---

[38] I explore this issue further in *Rights Reconsidered.*

NOTES AND NEWS

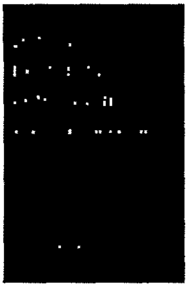
The editors are saddened to report the death of Richard Arthur Wollheim, on November 4, 2003, at the age of 80. Born in London in 1923, Professor Wollheim studied at the Westminster School, then at Balliol College, Oxford. He served in the infantry from 1942 to 1945; in 1944, he was briefly held as a prisoner of war and escaped. From 1949 until 1982 he taught philosophy at University College, London, where he was named Grote professor of mind and logic in 1963. In 1982, he moved to America, teaching first at Columbia University, then moving in 1985 to the University of California, Berkeley, where he was chair of the philosophy department from 1998 to 2002. Between 1989 and 1996 he split his time between Berkeley and the University of California at Davis.

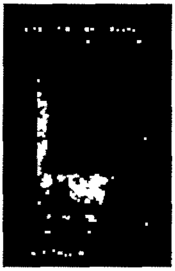
Wollheim received numerous awards and honors throughout his career; in recent years he was the Luce Visiting Scholar at Yale's Humanities Center, the Gareth Evans Memorial Lecturer at Oxford, the Roland Penrose Lecturer at London's Tate Gallery, the Werner Heisenberg lecturer at the Bavarian academy, the Lindley lecturer at the University of Kansas, and a visiting scholar at the J. Paul Getty Institute in Los Angeles. He was president of the Aristotelian Society from 1967–68, the British Society for Aesthetics from 1993, and the Pacific Division of the American Philosophical Association in 2002–03. He became a fellow of the British Academy in 1982 and of the American Academy of Arts and Sciences in 1986.

Among his best-known books are *Art and Its Objects* (1968, second edition, 1980); a collection of essays entitled *On Art and the Mind* (1973); his 1982 William James lectures at Harvard, published as *The Thread of Life* (1984); his 1984 A.W. Mellon Lectures in the Fine Arts at the National Gallery of Art, published as *Painting as an Art* (1987); *The Mind and Its Depths* (1993); the 1991 Cassirer lectures at Yale, published as *On the Emotions* (1999). His one novel, *A Family Romance*, was first published in 1969.

Wollheim was among the finest philosophers of the Anglo-American postwar era, embodying the intellectual virtues of analytic philosophy in the clarity of his writing and ingenuity of his arguments but engaged by subjects typically more central to continental thinking, in particular, aesthetics and the visual arts. In his work on personal identity and the philosophy of mind, Wollheim focused on psychoanalysis, regarding it the best kind of theory through which to understand human nature. He wrote with authority on Freud and Freudianism in general, particularly Melanie Klein. Psychoanalysis was crucial to his personal outlook and also played a fundamental role in defining his outlook on art, though he disclaimed any intention of psychoanalyzing works of art per se. *Art And Its Objects* contains what is widely regarded as Wollheim's major philosophical contribution, which he designated "seeing in." Pictorial perception is a complex achievement—what we see in a painting was intended by the artist, who organized the surface in order that viewers should grasp particular meanings. The viewer infers the intention from the way the painting looks, and this, Wollheim believed, "presupposes a universal human nature in which artist and audience share." Wollheim engaged in the life of art as a critic and an enthusiast, as well as through being a philosopher. He wrote widely and brilliantly about the artists he admired—painters in whose work the universal human nature, in which he believed, was palpably present. Whatever one may think of the detail in Wollheim's analyses and interpretations of art, he took a brave stand against the reductionisms that impoverish the way so many intellectuals have approached what are, in effect, the highest achievements of the human spirit.

New from
Oxford





Deflating Existential Consequence
A Case for Nominalism
JODY AZZOUNI
Escaping metaphysical straitjackets (such as the correspondence theory of truth), while retaining the insight that *some* truths are about objects that *do* exist, Azzouni says that we can sort scientifically-given objects into two categories: ones which exist, and to which we forge instrumental access in order to learn their properties, and ones which do not, that is, which are made up in exactly the same sense that fictional objects are.
2004 $49.95

New in Paperback!
The Animal Question
Why Non-Human Animals Deserve Human Rights
PAOLA CAVALIERI
Translated by CATHERINE WOOLLARD
"Ms. Cavalieri shows that contemporary discussions in ethics and bioethics risk arbitrariness or incoherence because they have failed to tackle the issue of the status of animals. From now on, opponents of animal rights must try to answer Ms. Cavalieri's argument, and anyone writing in bioethics will have to meet her challenge."—Peter Singer, Princeton University
paper 2004 $16.95

New in Paperback!
Purple Haze
The Puzzle of Consciousness
JOSEPH LEVINE
In this wide-ranging study, Joseph Levine explores both sides of the mind-body dilemma, presenting the first book-length treatment of his highly influential ideas on the "explanatory gap," the fact that we can't explain the nature of phenomenal experience in terms of its physical realization.
paper 2004 $19.95

New in Paperback!
Against Equality of Opportunity
MATT CAVANAGH
"This is a well-argued, insightful, highly-nuanced book, well worth close study by anyone who is concerned to make sense of the concept of equal opportunity."—Louis Pojman, *Notre Dame Philosophical Reviews*
paper 2003 $19.95

New in Paperback!
Objects and Persons
TRENTON MERRICKS
"With style and wit, Merricks stakes out an original position and defends it with an abundance of interesting, carefully formulated arguments. His book is sure to be one of the standard points of reference for the growing number of philosophers interested in these issues."—Michael B. Burke, *Philosophical Review*
paper 2003 $22.00

Prices are subject to change and apply only in the US. To order, or for more information, please call 1-800-451-7556
Visit our website at www.oup.com/us

OXFORD

CAMBRIDGE

Get into Focus

Charles Taylor
Edited by Ruth Abbey
Contemporary Philosophy in Focus
$60.00 Hardback 0-521-80136-2 232pp
$20.00 Paperback 0-521-80522-8

Art versus Non-Art
Art out of Mind
Tsion Avital
Contemporary Artists and their Critics
$85.00 Hardback 0-521-82465-6 470pp

Practical Conflicts
New Philosophical Essays
Edited by Peter Baumann and Monika Betzler
$70.00 Hardback 0-521-81271-2 352pp
$26.00 Paperback 0-521-01210-4

Ancient Anger
Perspectives from Homer to Galen
Edited by Susanna Morton Braund and Glenn W. Most
Yale Classical Studies 32
$65.00 Hardback 0-521-82625-X 336pp

John Stuart Mill
A Biography
Nicholas Capaldi
$40.00 Hardback 0-521-62024-4 464pp

The Mind and Its Stories
Narrative Universals and Human Emotion
Patrick Colm Hogan
Studies in Emotion and Social Interaction
$65.00 Hardback 0-521-82527-X 318pp

The Evolution of Reason
Logic as a Branch of Biology
William S. Cooper
Cambridge Studies in Philosophy and Biology
$29.00 Paperback 0-521-54025-9 236pp
Now in paperback

The Cambridge Companion to Thomas Reid
Edited by Terence Cuneo and René van Woudenberg
Cambridge Companions to Philosophy
$70.00 Hardback 0-521-81270-4 392pp
$26.00 Paperback 0-521-01208-2

Kant on Representation and Objectivity
A.B. Dickerson
$60.00 Hardback 0-521-83121-0 228pp

Self-Interest before Adam Smith
A Genealogy of Economic Science
Pierre Force
Ideas in Context 68
$65.00 Hardback 0-521-83060-5 296pp

The Cambridge Companion to Quine
Edited by Roger F. Gibson, Jr.
Cambridge Companions to Philosophy
$70.00 Hardback 0-521-63056-8 344pp
$25.00 Paperback 0-521-63949-2

Word and World
Practice and the Foundations of Language
Patricia Hanna and Bernard Harrison
$85.00 Hardback 0-521-82287-4 432pp
$30.00 Paperback 0-521-53744-4

Metaphor and Emotion
Language, Culture, and Body in Human Feeling
Zoltan Kovecses
Studies in Emotion and Social Interaction
$25.00 Paperback 0-521-54146-8 240pp
Now in paperback

www.cambridge.org
CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE

with Cambridge

Natural Kinds and Conceptual Change
Joseph LaPorte
Cambridge Studies in Philosophy and Biology
$70.00 Hardback 0-521-82599-7 237pp

A Physicalist Manifesto
Thoroughly Modern Materialism
Andrew Melnyk
Cambridge Studies in Philosophy
$70.00 Hardback 0-521-82711-6 340pp

Locke's Philosophy of Language
Walter R. Ott
$55.00 Hardback 0-521-83119-9 168pp

Hegel on Ethics and Politics
Edited by Robert B. Pippin and Otfried Hoffe
The German Philosophical Tradition
$65.00 Hardback 0-521-81814-1 360pp

Money and the Early Greek Mind
Homer, Philosophy, Tragedy
Richard Seaford
$70.00* Hardback 0-521-83228-4 c 550pp
$25.00* Paperback 0-521-53992-7

C.G. Jung and the Making of Modern Psychology
The Dream of a Science
Sonu Shamdasani
$28.00 Paperback 0-521-53909-9 404pp

States and Citizens
History, Theory, Prospects
Edited by Quentin Skinner and Bo Strath
$60.00 Hardback 0-521-83156-3 244pp
$22.00 Paperback 0-521-53926-9

The Stag Hunt and the Evolution of Social Structure
Brian Skyrms
$55.00 Hardback 0-521-82651-9 168pp
$20.00 Paperback 0-521-53392-9

The Evidence for the Top Quark
Objectivity and Bias in Collaborative Experimentation
Kent W. Staley
$70.00 Hardback 0-521-82710-8 360pp

A Systematic Theory of Argumentation
The Pragma-dialectical Approach
Frans H. van Eemeren and Rob Grootendorst
$58.00 Hardback 0-521-83075-3 224pp
$22.00 Paperback 0-521-53772-X

An Introduction to Mill's Utilitarian Ethics
Henry R. West
$55.00 Hardback 0-521-82832-5 232pp
$20.00 Paperback 0-521-53541-7

Descartes' *Meditations*
An Introduction
Catherine Wilson
Cambridge Introductions to Key Philosophical Texts
$55.00 Hardback 0-521-80981-9 284pp
$20.00 Paperback 0-521-00766-6

Lectures on Logic
Edited by J. Michael Young
The Cambridge Edition of the Works of Immanuel Kant
$40.00* Paperback 0-521-54691-5 c 727pp

Now in paperback

Prices subject to change

www.cambridge.org

CAMBRIDGE
UNIVERSITY PRESS

Graduate Faculty Philosophy Journal

The **Graduate Faculty Philosophy Journal**, published in association with the Department of Philosophy, New School for Social Research, is a forum for historical and contemporary issues in philosophy

VOL. 25 NO. 1

NATHAN ANDERSEN	Dynamic Boundaries: Place in Aristotle's Biology
GÉRARD GRANEL	The Thirties Are Still Before Us
JOSHUA L. KATES	Philosophy First, Last, and Counting
FRANÇOIS MATHERON	Althusser, or the Impure Purity of the Concept
JOHN NORAS	Sellars vs. Chisholm on Thinking, Introspection, and Language
THOMAS RENTSCH	Practical Certainty: Beyond Dogmatism and Relativism
SPIROS TEGOS	The Politics of Human Rights: A Politics of Pity?

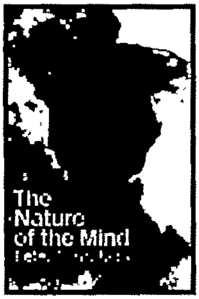
UPCOMING SPECIAL ISSUES

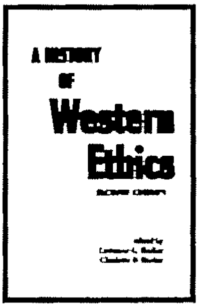
ESSAYS ON THE HISTORY AND PHILOSOPHY OF MATHEMATICS (25 2)

edited by Alexei Angelides

with contributions by Lanier Anderson, Juliette Kennedy, Peter Pesic, Danielle MacBeth, and others

All communications should be addressed to the Editor, **Graduate Faculty Philosophy Journal**, Department of Philosophy, New School for Social Research, 65 Fifth Avenue, New York, NY 10003. The *Journal* is biannual. Domestic rates: Individuals: $20.00/year; Students: $12.00/year; Institutions: $45.00/year.





Philosophy from Routledge

THE NATURE OF THE MIND
An Introduction
Peter Carruthers

"This book is excellent, certainly a deserving addition to the existing texts in the philosophy of mind. It is also distinctive both in the topics it covers and in its approach to those topics."
—Curtis Brown, Trinity University

320 pp • $24.95

WHAT WHITE LOOKS LIKE
African-American Philosophers on the Whiteness Question
Edited by George Yancy

"This remarkable and provocative collection brings together philosophy and activism, synthesis and critique, Marx and Fanon, feminism and anti-racism, humor and high seriousness."
—David Roediger, University of Illinois

304 pp • $24.95

Now in paperback
HISTORY OF JEWISH PHILOSOPHY
Edited by Daniel H. Frank and Oliver Leaman

This book explores the entire scope and variety of Jewish philosophy from philosophical interpretations of the Bible right up to contemporary Jewish feminist and postmodernist thought

952 pp • $47.95

A HISTORY OF WESTERN ETHICS
SECOND EDITION
Edited by Lawrence C. Becker and Charlotte B. Becker

This newly revised and updated edition of *A History of Western Ethics* is a coherent and accessible overview of the most important figures and influential ideas of the history of ethics in the Western philosophical tradition.

216 pp • $22.95

PERSONAL IDENTITY
SECOND EDITION
Harold Noonan

"Well informed and sophisticated discussion.... Very rewarding and challenging."
—*Mind*

In the second edition of *Personal Identity*, Harold Noonan presents the major historical theories of personal identity, particularly those of Locke, Leibniz, Butler, Reid and Hume.

256 pp • $31.95

THEORIES OF SCIENTIFIC PROGRESS
An Introduction
John Losee

Losee covers the interpretations of scientific progress from William Whewell through Karl Popper and Imre Lakatos to Thomas Kuhn and beyond, to the latest contemporary debates.

192 pp • $27.95

Routledge

1.800.634.7064 ◻ WWW.ROUTLEDGE-NY.COM

THE JOURNAL OF PHILOSOPHY

Celebrates 100 Years

Special Limited
Time Offer
Full Volumes 1-99
1904-2002

$50.00 per Volume

All volumes are available for $50.00
plus $5.00 shipping and handling.
Send check or money order to:

The Journal of Philosophy
1150 Amsterdam Avenue, MC 4972
Columbia University
New York, NY 10027
(212) 666-4419
www.journalofphilosophy.org

Calcutta University Library
ALIPORE CAMPUS

THE JOURNAL OF
PHILOSOPHY

VOLUME CI, NUMBER 3
MARCH 2004

page
111 *On the Systematicity of Language and Thought* Kent Johnson
140 *What Can Kant Teach Us about Emotions?* Maria Borges

159 NEW BOOKS

11/5/04

Published by **The Journal of Philosophy, Inc.**

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Akeel Bilgrami, Arthur C. Danto, Kent Greenawalt, Patricia Kitcher, Philip Kitcher, Isaac Levi, Mary Mothersill, Philip Pettit, Carol Rovane, Achille C. Varzi. *Editor Emeritus*: Sidney Morgenbesser. *Consulting Editors*: David Albert, John Collins, James T. Higginbotham, Charles D. Parsons, Wilfried Sieg. *Managing Editor*: John Smylie.

THE JOURNAL OF PHILOSOPHY is owned and published by the Journal of Philosophy, Inc. *President*, Arthur C. Danto; *Vice President*, Akeel Bilgrami; *Secretary*, Daniel Shapiro; *Treasurer*, Barbara Gimbel; *Other Trustees*: Lee Bollinger, Leigh S. Cauman, Kent Greenawalt, Michael J. Mooney, Lynn Nesbit.

All communications to the Editors and Trustees and all manuscripts may be sent to John Smylie, Managing Editor, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, New York 10027. FAX: (212) 932-3721.

You may also visit our website at: www.journalofphilosophy.org

THE JOURNAL OF PHILOSOPHY

2004

SUBSCRIPTIONS (12 issues)

Individuals	$35.00
Libraries and Institutions	$75.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$15.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are required.

Published monthly as of January 1977; typeset and printed by Capital City Press, Montpelier, VT.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL's terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

POSTMASTER: Periodical postage paid at New York, NY, and other mailing offices

POSTMASTER: Send address changes to the *Journal of Philosophy* at MC 4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027.

© Copyright 2004 by the Journal of Philosophy, Inc.

ISSN 0022-362X

THE JOURNAL OF PHILOSOPHY
VOLUME CI, NO. 3, MARCH 2004

ON THE SYSTEMATICITY OF LANGUAGE AND THOUGHT*

The last twenty years or so have seen a lot of discussion about systematicity—roughly the idea that if a person's repertoire contains the sentence/thought that John loves Mary, then her repertoire also contains the sentence/thought that Mary loves John.[1] Much of the discussion of systematicity has focused on the technical question of whether and how connectionists can explain the systematicity of cognitive capacities without simply implementing a classical architecture. That is, the debate has centered on whether connectionist systems offer a genuinely novel way to understand the empirical phenomenon of systematicity. Discussions of systematicity have not been confined to the connectionism/classicism debate. Systematicity (sometimes under the name "generality" or "recombinability") has also been discussed frequently in other areas of the philosophy of mind.[2] Two common threads unite these discussions of systematicity.

---

[*] I am grateful to Ernie Lepore, Robert May, and Greg Ray for much useful comments and criticism.

[1] For example, Jerry Fodor and Zenon Pylyshyn, "Connectionism and Cognitive Architecture: A Critical Analysis," *Cognition*, xxviii (1988): 3–71; Fodor and Brian P. McLaughlin, "Connectionism and the Problem of Systematicity: Why Smolensky's Solution Doesn't Work," *Cognition*, xxxv (1990): 183–204; Paul Smolensky, "The Constituent Structure of Mental States: A Reply to Fodor and Pylyshyn," *Southern Journal of Philosophy*, xxvi (1987): 137–60; Smolensky, "On the Proper Treatment of Connectionism," *Behavioral and Brain Sciences*, xi (1988): 1–23; Robert Hadley, "Systematicity in Connectionist Language Learning," *Mind and Language*, ix (1994): 247–72; Hadley, "Cognition, Systematicity, and Nomic Necessity," *Mind and Language*, xii (1997): 137–53; Lars Niklasson and Tim van Gelder, "On Being Systematically Connectionist," *Mind and Language*, ix (1994): 288–302; Robert Cummins, "Systematicity," this JOURNAL, xciii, 12 (December 1996): 591–614. The bibliographies of these papers contain many more citations

[2] For example, Gareth Evans, *The Varieties of Reference* (New York: Oxford, 1982), chapter 4; Charles Travis, "On Constraints of Generality," *Proceedings of the Aristotelian Society*, xciv (1994): 165–88; Christopher Peacocke, *A Study of Concepts* (Cambridge: MIT, 1992), chapter 2.

0022-362X/04/0103/111–39

First, claims about systematicity—especially the systematicity of language—are typically taken to be trivially true. For instance, Robert Matthews writes of "our capacity for natural language, which is by all accounts paradigmatically systematic."[3] Robert Cummins tells us that "the systematicity of language...is relatively unproblematic" (*op. cit.*, p. 591). Jerry Fodor and Zenon Pylyshyn assure us that "linguistic capacity is a paradigm of systematic cognition" (*op. cit.*, p. 37). In the case of thought, Christopher Peacocke writes, "The recombinability of concepts to form new thoughts has been largely unquestioned in the published literature" (*op. cit.*, p. 42). (He then goes on to assume this form of systematicity, and to develop an explanation for it.) Second, the nature of systematicity is rarely more than cursorily sketched: an example or two is typically thought to characterize the phenomenon adequately. In short, the prevailing optimism of the literature suggests that everybody knows what systematicity is, and that language (and probably thought, too) clearly has it.

Despite this widespread enthusiasm, I am skeptical about both these claims. I do not think it is very clear what systematicity is, and I am even more skeptical that language (or thought) has it. Since it is easier to study overtly realized phenomena like language, I will focus on the systematicity of language. I argue that as systematicity is typically conceived of in the literature, natural languages are not systematic. More importantly, I contend that this way of thinking of systematicity rests on a crucial but inaccurate assumption about the nature of linguistic natural kinds. When this assumption is corrected, it becomes utterly mysterious what it even means to say that natural language is systematic. Although my primary goal is to explore the nature of the systematicity of language as such, I also maintain that this exploration has consequences for the systematicity of thought and for the role that systematicity plays in the connectionism/classicism debate. The systematicity of thought, I argue, suffers from problems analogous to the systematicity of language, and it does not appear to play any interesting role in the debates about cognitive architecture. In sum, it would only be a slight exaggeration to say that my view of systematicity is maximally pessimistic: it does not exist; but if it did, languages would not have it; and if it existed and languages had it, systematicity still would not be all that interesting to many of the debates in which it has figured.

The article is structured as follows. I begin by developing a natural and intuitive characterization of the claim that language is systematic. In section II, I show that this claim is false. In section III, I defuse an objection to the argument against the systematicity of language by showing that the argument raised in section II is quite general and robust. In section IV, I increase this generality and robustness by illustrating another dimension of problems arising for a variety of natural and intuitive formulations of systematicity. In section V, I argue that the systematicity of thought is at least as problematic as the systematicity of language. In section VI, I briefly explore the role of systematicity in the connectionism/classicism debate. I conclude in section VII.

---

[3] "Can Connectionists Explain Systematicity?" *Mind and Language*, XII, 2 (1997): 154–77, here p. 175

I. AN OVERVIEW OF SYSTEMATICITY

In order to evaluate the claim that language is systematic, we need to better understand the claim. Unfortunately, most characterizations of systematicity are not very precise.[4] Of the very few explicit characterizations of systematicity, Robert Cummins offers one that gets really close to the heart of the matter:

> A system is said to exhibit systematicity if, whenever it can process a sentence *s*, it can process systematic variants of *s*, where systematic variation is understood in terms of permuting constituents or (more strongly) substituting constituents of the same grammatical category (*op. cit.*, p. 594).

I will adopt a version of Cummins's characterization, with a couple of minor qualifications. First, instead of speaking about a system "processing" sentences, I will speak about whether a sentence is grammatical or not. I do this primarily because I want to focus on whether language is systematic, and our linguistic abilities may be distinct from our abilities to process language. For example, although speakers reject **the child seems sleeping* as ungrammatical, there is also a sense in which it has a clear meaning, and might be thought to be processible, in some extended sense of the word.[5] Indeed, utterances like **him give she cookie now!* and *take* and *slab!* might also be processible, too. I wish to treat all these examples as falling outside of natural language—or at least outside of English, which is the natural language

---

[4] Cf. Niklasson and van Gelder, Cummins; and Hadley for similar comments to this effect.

[5] Throughout this article, I use a number of standard conventions (i) For ease of readability, I use italics instead of quotation marks when mentioning an expression. (ii) The asterisk denotes ungrammaticality. (iii) The term *grammar* denotes all the components of a natural language, including its syntax, semantics, and phonology. (iv) I use such terms as *kind* and *category* interchangeably.

I will focus on. (I suspect that I have not really adjusted Cummins's idea, since the quotation above might be referring to the system that constitutes our linguistic abilities.)

Second, for expository convenience, I will adopt what Cummins calls the "stronger" version of systematicity. According to this version, what is relevant to systematicity is not whether we can interchange *John* and *Mary* in *John saw Mary* and still have a grammatical sentence, but whether we can substitute in *any* words of the same grammatical category as *John* and *Mary*, and still preserve grammaticality. Since there are only finitely many primitives in our language, and sentences can be arbitrarily finitely long, it is not clear that this version is really stronger than the other.[6] In any case, nothing essential rides on adopting this version. All the points could be made with the other version simply by using sentences with, for example, some additional relative clauses, conjunctions, or the like.

With these qualifications made, we can now rearrange Cummins's criterion into the following characterization of systematicity. A language *L* is systematic if and only if (S) holds for all *A*:

> (S) *A* is a constituent of *L* only if for all *B* of the same linguistic kind as *A*, and all things *C*, *C* can compose with *A* (in a certain way) to form a sentence if and only if *C* can compose with *B* (in that same way) to form a sentence.

(S) also captures a type of systematicity that has been discussed outside of the connectionism/classicism debate. Gareth Evans, for instance, discusses what he calls the "Generality Constraint":

> It seems to me that there must be a sense in which thoughts are structured. The thought that John is happy has something in common with the thought that Harry is happy, and the thought that John is happy has something in common with the thought that John is sad (*op. cit.*, p. 100).

> The Generality Constraint requires us to see the thought that a is F as lying at the intersection of two series of thoughts: the thoughts that a is F, that a is G, that a is H,..., on the one hand, and the thoughts that a is F, that b is F, that c is F,..., on the other (*op. cit.*, p. 209; cf. p. 104, fn. 21).

Although Evans is primarily concerned with thought, he illustrates the Generality Constraint with language (*op. cit.*, pp. 101–03). Others have discussed it in a linguistic context.[7] Thus, (S) captures a property that is often believed to hold of language.

---

[6] There are, however, infinitely many words in the typical speaker's repertoire, because there are recursive word-forming devices, for example, *great-grandmother, great-great-grandmother, great-great-great-grandmother*.

Before moving on, notice that (S) is essentially a condition on the individuation of linguistic kinds. (I use such terms as *linguistic kind, grammatical category,* and so forth, interchangeably.) For example, (S) is equivalent to (1):

> (1) If *A* but not *B* can compose with *C* in a certain way to form a sentence (of *L*), then *A* and *B* do not belong to the same linguistic kind (of *L*).

Although I contend that there is a serious flaw with (S)—and the versions of systematicity like Cummins's from which (S) is extracted—there is nevertheless something importantly right about (S). (S) correctly emphasizes that any statement of systematicity is essentially relativized to a theory of linguistic kinds. After all, language is supposed to be systematic because when there are sentences like *John sings* and *Mary jumps,* there are also sentences like *Mary sings* and *John jumps.* We do not want to say that language is not systematic simply because we do not also have **Jumps sings,* where a verb is substituted for a noun. The importance of a theory of kinds of constituents to systematicity would seem obvious. However, sometimes it goes unmentioned (for example, Fodor and Brian McLaughlin's well-known discussion of the systematicity of thought never mentions that the elements that systematicity predicts can be legitimately switched around must be of the same kind (*op. cit.*)). Although systematicity's dependence on a theory of linguistic (or mental) kinds may be too obvious to bear mention, I hold that this dependence renders standard formulations of systematicity like (S) both false and otherwise unacceptable. My first goal is to show that there is no obvious sense in which (S) is true.

II. THE ARGUMENT AGAINST THE SYSTEMATICITY OF LANGUAGE

We have just seen that (S)—and indeed any version of the systematicity of natural language—presupposes some theory of linguistic kinds of expressions. So before we can explore (S) we need to fix on some theory of kinds of expressions. Following common trends in philosophy and linguistics, we might take the relevant kinds to include verbs (transitive and intransitive), quantifiers, connectives, adjectives, nouns, and singular terms. I will work with these kinds of expressions mainly because they are the ones that get discussed in this literature, when kinds are mentioned at all. (In the long run, the particular kinds chosen will not matter. In the next section, I will argue that the lack of systematicity exhibited with these kinds is actually a weak instance of a more serious problem that holds for any realistic choice of linguistic kinds of expressions.) Given some collection of linguistic kinds, (S) predicts that we cannot find a pair of terms *A* and *B* belonging to one of these kinds and a linguistic environment *C* such that the composition *C*(*A*) is grammatical but *C*(*B*) is not. For the linguistic kinds mentioned above, this prediction is false. Below I provide counterexamples for each of these categories. It bears emphasis, though, that the nonsystematicity of natural language is a *pervasive* phenomenon: many more instances of the sorts of examples I offer can be found throughout the linguistics literature.[8]

---

[7] For example, Travis.

I begin with verbs. I will dwell a bit on this example, in order to display some of the issues involved, and I will then move more quickly through the other kinds of words. Consider the verbs *put* and *stow*. On the one hand, *put* appears in the structure given in (2a), but not in the structure given in (2b). In contrast, the distribution of *stow* is precisely the opposite: it appears in the second structure but not the first, as (2c–d) show:

(2) a. John put his gear down.
b. *John put his gear.
c. *John stowed his gear down.
d. John stowed his gear.

Even though you can neither put nor stow something without putting or stowing it somewhere, only the verb *put* demands that some indication be given as to where the item was placed. The distribution in (2) appears to be due, in large part at least, to syntactic properties of the word. However, the grammatical behavior of a given word is influenced not only by its (morpho-) syntactic properties, but also by its semantic and phonological properties. For instance, consider the well-known "locative alternation."[9] Words like *drench* and *throw* exhibit complimentary distribution over two grammatical forms:

---

[8] It is also worth observing that many languages make many other distinctions that are not found in English, so some languages contain many striking failures of systematicity over these linguistic kinds.

[9] For example, James Higginbotham, "Elucidations of Meaning," *Linguistics and Philosophy*, XII (1989): 465–517; Steven Pinker, *Learnability and Cognition* (Cambridge: MIT, 1989); Jess Gropen, Pinker, Michelle Hollander, and Richard Goldberg, "Syntax and Semantics in the Acquisition of Locative Verbs," *Journal of Child Language*, XVIII (1991): 115–51.

(3) a. I drenched the flowers with water.
b. *I drenched water onto the flowers.
(4) a. *I threw the flowers with water.
b I threw water onto the flowers.

(The interpretations of (4a) on which it is grammatical are not relevant here.) It is common to maintain that *drench* does not fit into the second form because the meaning of *drench* does not specify how the water got onto the flowers, but only specifies the end state (being drenched) of the flowers. In contrast, *throw* does say something about how the water got onto the flowers, but it does not specify what happened to the flowers as a result of throwing water on them (for example, they may or may not have been drenched). Here, then, is a case where the semantics appears to affect the grammatical properties of some words. In the case of phonology, consider the two verbs *show* and *describe*. They appear to be syntactically and semantically very similar. However, if you wanted to say that Alice showed or described the book to Martha, only in the case of *show* would you use the double-object construction:

(5) a. Alice showed Martha the book.
b. *Alice described Martha the book.

(5a–b) differ because English places phonological restrictions on what verbs can enter into the double-object construction. Roughly speaking, the differing behavior in (5) is due to the fact that *show* but not *describe* is typically stressed on its first syllable. This is seen elsewhere with, for example, *I told Martha a story* versus **I reported the police the accident*.[10]

Let us move on to the remaining categories. Quantifiers differ widely in their grammatical behavior, even when they are apparently similar in meaning. For instance, *each* and *every* both seem to express universal quantification, but they are not systematic in distribution:[11]

(6) a. *The men every read the book carefully.
b. The men each read the book carefully.
c. One boy ate almost every apple in the basket.
d. *One boy ate almost each apple in the basket.

Similarly, the English words *and* and *but* are often thought to be extremely similar. In fact, it is commonly supposed that the two differ only in their pragmatic implicatures.[12] However, their differences are linguistically deeper.

---

[10] Cf., for example, Pinker, pp 45–47 for more discussion.

[11] For example, Filippo Beghelli and Tim Stowell, "Distributivity and Negation: The Syntax of *Each* and *Every*," in Anna Szabolcsi, ed., *Ways of Scope Taking* (Boston: Kluwer, 1997), pp. 71–107.

(7) a. John and Mary carried the piano upstairs together.
b. *John but Mary carried the piano upstairs together.
c. But for Kristen's help, I would have failed the test.
d. *And for Kristen's help, I would have failed the test.

Other predicates like adjectives can also have restrictions on where they appear. Suppose the basket is full of apples, and John is fond of pie. Our language will allow you to say that the basket is full without specifying what it is full of. Not so for fondness:

(8) a. The basket is full.
b. *John is fond.

As for nouns, we have, for example, the dissimilar distribution given in (9).

(9) a. McDonald's destruction by Burger King had a major impact on beef production.
b. *McDonald's resemblance by Burger King had a major impact on beef production (cf. *McDonald's resemblance to Burger King*).[13]

Our final category is the singular terms, which encompass such things as proper names, pronouns, and demonstratives. Singular terms are often heralded as the paradigms of systematicity: since *John loves Mary* is a sentence, so is *Mary loves John*. Strictly speaking, the two occurrences of *John* (and *Mary*) are different in some respects. Although English typically does not give phonological expression to its case assignment, a great many languages do overtly distinguish between nominative and accusative case. In Attic Greek, for instance, *Socrates loves Xanthippe* is expressed by (10a), and *Xanthippe loves Socrates* is expressed by (10b):

(10) a. ho Socratês philei tên Xanthippên.
b. hê Xanthippê philei ton Socratên.

Moreover, there is ample reason for supposing that case assignment is a fundamental property of all languages. In English, case assignment is overt in pronouns: *He loves her* is a sentence and so is *She loves him*, but neither **Him loves she* nor **Her loves he* are. I will not dwell on the facts about case assignment, although they are surely relevant to a precise formulation of the systematicity of language. A further look at the behavior of singular terms shows that they exhibit other different properties. For instance, it is possible to treat names as common nouns and quantify over them. However, the same cannot be done with other singular terms. If you wanted to announce that two women in your class, both of whom were named *Mary*, needed to report to the principal's office, you might use a sentence like (11a). However, even if the two women were standing near you so that you could easily demonstrate them, you would never use (11b) or (11c).

---

[12] For example, H.P. Grice, *Studies in the Ways of Words* (Cambridge: Harvard, 1989), p. 46.

[13] For example, Carol Tenny, *Aspectual Roles and the Syntax-Semantics Interface* (Boston: Kluwer, 1994).

(11) a. The two Marys should report to the principal's office immediately.
b. *The two shes should report to the principal's office immediately.
c. *The two these should report to the principal's office immediately. (Observe that the subject is distinct from the partitive construction *the two of these*; compare **the two of Marys*, **the two of shes*.)[14]

Furthermore, pronouns and demonstratives differ from one another in that only the former has a possessive form. Curiously, complex demonstratives can take the possessive:[15]

(12) a. Katie's leg is broken / Bob and Susie's legs are broken.
b. His leg is broken / Their legs are broken.
c. *This's leg is broken / *Those's legs are broken.
d. This dog's leg is broken / These dogs' legs are broken.

As an aside, it is worth noting that the failure of systematicity extends beyond individual words to complex constituents of sentences. It has been suggested for instance, that the direct object of *believes* is the same in both of (13a–b):[16]

(13) a. John believes that snow is white.
b. John believes the proposition that snow is white.

There are deep and difficult issues concerning the identity of the meaning of the content clause of propositional attitude reports. However, regardless of these issues, human languages do not treat the clause *that* S and the noun phrase *the proposition that* S alike. For instance, although both expressions are fine in (14), only *that* S can be moved to the end of the sentence in (15):

---

[14] Note that even though the latter two sentences are ungrammatical, they still have a clear meaning. But many sentences have a meaning even though there is something structurally wrong with them: **The child seems sleeping* is one famous case of this.

[15] Kent Johnson and Ernie Lepore, "Does Syntax Reveal Semantics? A Case Study of Complex Demonstratives," in James E. Tomberlin, ed., *Philosophical Perspectives: Language and Mind*, xvi (2002): 17–41.

[16] Stephen Schiffer, "Meanings," in Joseph Keim Campbell, Michael O'Rourke and David Shier, eds., *Meaning and Truth* (New York: Seven Bridges, 2002), pp. 79–102.

(14) a. That snow is white is commonly known.
b. The proposition that snow is white is commonly known.

(15) a. It is commonly known that snow is white.
b. *It is commonly known the proposition that snow is white.[17]

There are numerous other such differences between the two claims. Although the town gossip might speak of *the guy who Jane believes that Jill likes*, nobody speaks of **the guy who Jane believes the proposition that Jill likes*. Indeed, language's resistance to systematicity between these two forms of content clauses is so strong that even the presence or absence of *that* in a propositional attitude report can matter. It was observed long ago[18] that if the subject of a *that*-clause is pulled out of its normal place, as happens when we form questions, the complementizer *that* is better left out of the sentence. For instance, **Which professor do you hope that will teach the course?* reliably jars speakers when they encounter the word *that*, but remove the *that* and the sentence is fine. These differences between expressions of the form *that* S and *the proposition that* S (and even S alone) underscore the complex differences between individual elements of language, a difference that (S) might appear to deny.

The failure of systematic behavior of expressions of natural language was noted early on in the linguistics literature, even before linguistics was understood as a branch of cognitive science. According to Zelig Harris, words should be individuated solely by their distributional properties.[19] That is, Harris held that natural language lacks systematicity so much that A and B are distinct words if and only if there is a C such that $C(A)$ is grammatical but $C(B)$ is not. It is probably an exaggeration to say no two words pattern in exactly the same way. However, the fact that some linguists have found this view plausible shows just how differently seemingly similar words behave. In the next section, I consider an attempt to restore the systematicity of language despite this high degree of variability across words.

---

[17] For more discussion, cf. Noam Chomsky, *Knowledge of Language* (Westport, CT: Praeger, 1986), ch. 3

[18] J.R. Ross, "Constraints on Variables in Syntax" (Ph.D. diss., MIT, 1967).

[19] *Methods in Structural Linguistics* (Chicago. University Press, 1951).

III. THE POSSIBILITY OF VERY FINE-GRAINED LINGUISTIC CATEGORIES

We have just observed that if we take seriously the kinds of grammatical categories that are typically appealed to, then natural language is not systematic. At this point, though, it would be natural to object that the evidence just reviewed shows only that we have not found the right linguistic categories to use in a statement of systematicity. Let me put this objection in its strongest form.

> Look, there is an a priori reason why your argument cannot work. You have presented numerous linguistic differences between ostensibly similar words, but it does not follow from this that you have undermined the systematicity of language. We can always interpret these differences as showing only that the relevant linguistic types must be more fine grained than appears at first glance. That is, whenever you come up with some new linguistic details, we simply create more fine-grained taxonomies of linguistic types in such a way that systematicity is preserved. The idea is that if you show that *A* and *B* in category *X* behave differently, then we will just break *X* into two types: the *Xs-that-behave-like-A* and the *Xs-that-behave-like-B*. This strategy is of course completely general. So the fact that two words are grammatically different is evidence that the words are in different categories, and that language is systematic over these fine-grained categories.

Since the objection just given is in essence a proposal to use very Fine-Grained linguistic Categories, I will refer to it as FGC. If FGC is right, then nothing I have said so far is relevant to the issue of the systematicity of language. Moreover, FGC is driven only by a straightforward interpretation of systematicity. In particular, (S) says that if two words behave grammatically differently, then they belong to different linguistic kinds. FGC operates by assuming that the appropriate taxonomy of linguistic natural kinds is a *partition* over the set of words: each word belongs to one and only one type. (A partition of a set *X*, recall, is a set of subsets of *X* such that each element of *X* belongs to exactly one of these subsets.) Since each set in this partition is by definition a set of grammatically equivalent expressions, FGC guarantees systematicity-preserving categories. Furthermore, FGC's strategy appears to be completely innocent. Surely the grammatical types are what are responsible for differing grammatical behaviors. So if two words have different linguistic properties, then they must be of distinct linguistic types.[20] What could possibly be wrong with FGC?

---

[20] Such a position appears to have been adopted by W V. Quine, *Philosophy of Logic* (Englewood Cliffs, NJ: Prentice-Hall, 1970), p. 18.

FGC appears very compelling at first glance. However, it contains an important methodological flaw: it assumes that all the relevant kinds of expressions for its theory of systematicity are natural kinds. This looks like a small point, but it is crucial. FGC assumes that if *X* and *Y* are legitimate natural kinds of expressions, there is also a third natural kind *X-and-Y*, whose extension is the intersection of the extensions of *X* and *Y*. But it is a principle of the philosophy of science that natural kinds do not in general distribute over Boolean combinations of natural kinds. For instance, a gold sphere has metallurgical properties in virtue of being gold and mechanical properties in virtue of being spherical. But there is no natural kind gold-sphere that is employed in any of the sciences to explain the behavior of the object. A precisely similar case, I argue, holds in the organization of language. If linguistic natural kinds are not closed under Boolean combinations, then the fine-grained categories that FGC relies on will not be available, and FGC is therefore untenable as a strategy for preserving systematicity.

To begin, notice that the kinds referred to in (S)—or any other statement of systematicity—are linguistic natural kinds of expressions. The reason for this is, of course, that some kinds of words are not relevant, and if we attended to all possible kinds of words, putatively empirical linguistic generalizations involving reference to kinds of words become trivial and unexplanatory. In the cases of (S) and FGC, the claims become trivially false or vacuous, because the assumption that each word belongs to a unique natural kind is trivially false. So obviously, only some kinds of words are relevant. But as Goodman showed, unless we can say which kinds are relevant, the basic problem remains. There is, for instance, a class of words that Sally likes, and a class of words that she does not like (Sally is neutral about all possible words that she does not know or that have not yet entered the language, so she does not like them). Does the existence of these classes guarantee that English is "Sally-systematic" in the following sense?

(16) *A* is a constituent of language only if for all *B* of the same Sally-kind as *A*, and all things *C*, *C* can compose with *A* (in a certain way) to form a sentence that contains a word that Sally likes if and only if *C* can compose with *B* (in that way) to form a sentence that contains a word that Sally likes.

Given how Sally-kinds are defined, (16) is trivially true, but uninteresting. It would be interesting only if there were some independent characterization of the Sally-kinds, that is, if the Sally-kinds did some theoretical work beyond securing the truth of (16). For analogous reasons, the relevant kinds to be used in a statement of systematicity cannot just be stipulated to be those whose extensions are always sets of grammatically equivalent words. In such a case, the claim that language is systematic is a tautology, epistemically trivial, and scientifically useless. So in forming a statement of systematicity, we need to appeal to those kinds of words that really do have some independently useful status. To that end, we should look to linguistics to see what kinds of words are actually used in theorizing. If language turns out to be systematic over these kinds—or over some subset of them that we can argue are the relevant ones—then language is systematic; if it is not systematic over these kinds, language is not systematic. (That is, language may not be systematic if the complex *interactions* between linguistic natural kinds account for the relevant linguistic facts without rendering generalizations like (S) true.) Demanding that we use the kinds actually posited by linguistics restores systematicity to its rightful status as an empirical claim. It also follows that FGC now has empirical content: it predicts that linguistic theorizing will produce the fine-grained linguistic natural kinds that FGC needs. I argue that it does not.

To start the argument against the fine-grained categories that FGC needs, let me introduce two kinds of verbs.[21] The first kind of verb is the causatives: for example, *break, snap, soften, warp,* and so forth. Causatives all express that the subject of the verb brought it about that the object of the verb entered into some state. For example, *X break Y*—as in *Susan broke the vase*—has roughly the meaning *X* cause *Y* to be broken. There are numerous reasons for thinking that causatives form a natural kind of verb, and that verbs of this kind are distinctive in possessing a bit of structure in the verb's linguistic representation that means roughly "cause."[22] One reason for distinguishing the causative verbs is that (ceteris paribus) they have an intransitive form in which the subject of the transitive verb disappears and the object assumes the subject position, as in *the vase broke.* The second type of verb is the object-deletion verbs, for example, *eat, sew,* and *write.* Object-deletion verbs are in a way the converse of causatives. Although they are typically transitive, they can also appear as intransitive verbs with their object understood, as in *While Mary was eating/sewing/writing, the phone rang*. Importantly, some verbs (for example, *boil*) are causatives but not object-deletion verbs, other verbs (for example, *eat*) are not causatives but are object-deletion verbs, and yet further verbs (for example, *cook*) are both. These facts are witnessed in (17)–(19):

---

[21] The two kinds of verbs I use are for illustrative purposes only. You may, if you like, replace them with other interacting types of expressions.

[22] For example, Pinker; Terence Parsons, *Events in the Semantics of English* (Cambridge: MIT, 1990); Lisa Travis, "Event Structure in Syntax," in Carol Tenny and James Pustejovsky, eds., *Events as Grammatical Objects* (Stanford: CSLI, 2000), pp. 145–85. This bit of structure is overtly realized as a morpheme in hundreds of languages; cf. Kenneth Hale and Samuel Jay Keyser, "On Argument Structure and the Lexical Expression of Syntactic Relations," in Hale and Keyser, eds., *The View from Building 20* (Cambridge: MIT, 1993), pp. 53–109, see especially p. 102.

(17) a. Mary boiled the meat.
b. The meat boiled.
c. *Mary boiled.
(18) a. Mary ate the meat.
b. *The meat ate.
c. Mary ate.
(19) a. Mary cooked the meat.
b. The meat cooked.
c. Mary cooked.

(The readings of the ungrammatical sentences on which they are grammatical are not relevant here.) How should we handle the distribution in (17)–(19)? For proponents of FGC, the answer is clear. (17)–(19), according to FGC, is evidence that there are three linguistic categories instead of two. That is, there must be a category X of causative words like *boil* which are not also object-deletion verbs, a category Y of object-deletion verbs like *eat* which are not causatives, and a category Z of causative object-deletion verbs like *cook*. Of course, further linguistic considerations will break each of X, Y, and Z into multitudes of even more fine-grained categories $X_1,\ldots, X_m$, $Y_1,\ldots, Y_n$, $Z_1,\ldots, Z_k$, but I will not worry about these further subdivisions here. We need only observe how the FGC strategy works in a particular case.

The problem with FGC should now be clear. The linguistic evidence much more strongly supports the claim that there are two overlapping categories—the causatives and the object-deletion verbs—than it supports the claim that there are the three categories X, Y, and Z. The categories X, Y, and Z do nothing whatsoever to explain why X-type verbs and Z-type verbs share some of their grammatical behavior, and Y- and Z-type verbs share other grammatical behavior. As far as X, Y, and Z are concerned, the similarities between *boil* and *cook* and *eat* and *cook* are utterly mysterious. The similarities of behavior just mentioned are easily explained by a theory that includes causatives and object-deletion verbs among its linguistic kinds: *boil* and *cook* share certain grammatical behaviors because they are both causatives, and *eat* and *cook* share other behaviors because they are both object-deletion verbs. So it is plausible that a linguistic theory will include at least the categories of causatives and object-deletion verbs. But could the theory also contain the categories *X*, *Y*, and *Z*? Why would it? In general, anything that *X*, *Y*, and/or *Z* can explain is also explicable using the categories of causatives and object-deletion verbs. That is, any laws of linguistics statable with the categories *X*, *Y*, and/or *Z* is statable using the categories of causatives and object-deletion verbs. But if *X*, *Y*, and *Z* do not figure into any laws—linguistic, psychological, or otherwise—then that is good reason to hold that they are not real natural kinds. It is of course an empirical claim that *X*, *Y*, and *Z* are not needed in a successful linguistic theory. The total pattern of linguistic data could turn out so that overall it is vastly easier to use *X*, *Y*, and *Z* instead of causatives and object-deletion verbs. But there appears to be no evidence for this outcome, and there certainly is not enough evidence to make it even slightly plausible that *X*, *Y*, and *Z* are genuine linguistic categories. Barring further evidence, we should conclude that *X*, *Y*, and *Z* are not linguistic natural kinds.[23] With the loss of *X*, *Y*, and *Z*, the kinds FGC requires in order to account for (17)–(19) are simply not available. So FGC cannot be maintained as a viable linguistic theory, and so FGC fails as an attempt to preserve the systematicity of language.[24]

The argument just given, which was based around (17)–(19), should not be surprising. After all, in the previous section, we saw that the distribution of an expression can be affected not only by what (morpho-) syntactic natural kinds the expression belongs to, but also by its membership in various semantic and phonological kinds. Although there are various famous points of interaction between these areas, it also remains true that syntax, semantics, and phonology are to some extent independent of one another. The work of practicing phonologists, for instance, is often isolated from semantics and syntax. Thus, it is to be expected that many words can have a given phonological, semantic, or syntactic property independently of many of the other properties they might have in the remaining two linguistic domains. When such properties are relevant to the grammatical behavior of the expression, we will have yet further instances of overlapping linguistic natural kinds of the sort seen in (17)–(19).

---

[23] I stress that natural kinds, whether of linguistics, psychology, physics, or any other of the sciences, are determined by the best available scientific theories of the subject. Someone who wants to contend that *A*, *B*, and/or *C* are natural kinds needs to show why they are plausibly needed by some leading scientific theory.

[24] It is worth noting that determining the correct linguistic categories is an ongoing project in contemporary linguistics. Thus, it is simply false to say, as Fodor does, that "[i]nsofar as there are 'theory neutral' data to constrain our speculations about language, [systematicity] surely ought to count as one of them"—*Psychosemantics* (Cambridge: MIT, 1987), p. 150. Even saying what systematicity is, much less whether language has systematicity or not, rests on the highly theoretical notion of a taxonomy of linguistic kinds over which systematicity is defined.

IV. WHAT IS SYSTEMATICITY?

The argument of the preceding section exposes a fairly general problem for theories of systematicity. FGC is false, we have seen, because it attempts to enforce a partition of all the words of the language. However, the empirical facts strongly suggest that the linguistic natural kinds of words do not form anything as neat as a partition; rather, they form a much more complex overlapping structure. A given word belongs to multiple overlapping linguistic kinds, and no one of these kinds provides a total account of the word's grammatical behavior. If this is right, then (S) is probably false, because it implies that a word belongs to exactly one linguistic natural kind. One might try to argue that there is only one linguistic kind per word that is relevant to systematicity. But such a position would be extremely hard to defend, since it would amount to something like the claim that the difference between *John ate* and **John boiled* is relevant to systematicity, but the difference between **the meat ate* and *the meat boiled* is not. In such a case, it would be hard to see what systematicity even amounts to, much less why it matters.

At this point, a likely suggestion is that although (S) is false, a better statement of systematicity can be had. One might suggest that a language *L* is systematic if and only if (S′) holds for all *A*:

> (S′) *A* is a constituent of *L* only if for all *B* of *all* the same linguistic kinds as *A*, and all things *C*, *C* can compose with *A* (in a certain way) to form a sentence if and only if *C* can compose with *B* (in that way) to form a sentence.

The problem with (S′) is that it is a tautology. The sum total set of linguistic kinds that a given word belongs to exhausts the word's grammatical properties, by the very definition of the linguistic kinds. So if *A* is a constituent of language, then it is trivially true that any *B* of all the same linguistic kinds as *A* will behave grammatically identically to *A*, and so a fortiori *A* composes with *C* to form a sentence if and only if *B* does too. But systematicity is not supposed to be a triviality. Systematicity is supposed to be an empirical property of human languages, one which is sometimes thought to be easily explicable by recursive, symbol-processing computational devices, but which is not so easily explained by connectionist devices that do not simply implement the former sort of machinery.[25] If systematicity was a trivial property of language (or cognition), then there would be no need for any linguistic (or cognitive) theory to explain it. The worst that a theory or model could do would be to employ the wrong linguistic (or mental) kinds in its implementation of language (or cognition). But this sort of error would only be to get some of the empirical facts wrong, which is only a normal occurrence in the progression of any ongoing empirical science. So whatever notion of systematicity cognitive scientists worry about, it is not (S′). I will return to this point below.[26]

---

[25] For example, Fodor, p. 153; Fodor and Pylyshyn; Fodor and McLaughlin, Niklasson and van Gelder, Smolensky, "Constituent Structure and Explanation in an Inte-

Here is where we have gotten. The really deep problem with a standard characterization of systematicity like (S) is that it makes the false (or at least unjustified) assumption that each constituent of language belongs to exactly one relevant category. Moreover, systematicity becomes trivial if we demand only that it apply only to constituents that belong to all of the same categories. At this point, the following thought might arise. Since the linguistic environments a given word can grammatically appear in is determined by the word's grammatical categories, perhaps we can adjust the characterization of systematicity so that whether systematicity requires two words to be intersubstitutable depends not only on what linguistic kind(s) the words belong to, but also on what kind of linguistic environment we are dealing with. We might then entertain something like:

(S″) If *A* and *B* belong to category *X*, and *C* is an *X*-type linguistic environment, then *C*(*A*) is grammatical if and only if *C*(*B*) is too.

Thus, we might say that the environment *John* ___ *his gear under the bed* is a kind of transitive verb environment, and *put* and *stow* are transitive verbs, but *John* ___ *his gear* is a location-unspecified environment, and *stow* but not *put* is this type of transitive verb. Could this sort of strategy render language systematic? There are several problems with (S″). For starters, it suffers some by now well-known problems. First, (S″) is just a triviality, saying little more than language is systematic except when it is not. Second, it is not obvious that all of the requisite kinds *X* actually exist. Many linguistic environments require that an expression be contained in multiple categories, where the intersection of these categories is not a linguistic natural kind, which would make it hard to specify what type of linguistic environ-

---

grated Connectionist/Symbolic Cognitive Architecture," in Cynthia Macdonald and Graham Macdonald, eds., *Connectionism: Debates on Psychological Explanation* (Cambridge: Blackwell, 1995), pp. 223–90.

[26] Note also that this general problem of making systematicity trivial also applies against the FGC approach, if the linguistic kinds FGC requires actually existed.

ment it is. Similarly, using the categories of linguistic theory, (S'') is false. Returning to (20) (repeated below) above, linguistic theory correctly predicts that both *show* and *describe* are of the same syntactic category, and so both should fit into the double object construction:

(20) a. Alice showed Martha the book.
b. *Alice described Martha the book.

The fact that *describe* does not fit into this construction is not due to the fact that it *lacks* membership in any relevant category, but that it is *also* a member of a certain phonological category, which prevents it from entering into certain constructions. Thus, being a member of a certain category can prevent an expression from entering into a construction that it would otherwise normally participate in. The grammatical behavior of a word is not determined merely by what categories it belongs to, but also by what categories it does not belong to.

At this point, it would be ideal to unveil a victorious characterization of systematicity. I would like to replace (S) (and (S') and (S'')) with something that is responsive to the linguistic facts and which has all the bite needed to maintain current discussions involving systematicity. Unfortunately, I do not have such a characterization. However, our discussion so far provides some clues as to what is needed from a useful notion of the systematicity of language. I conclude this section with four desiderata on a theory of systematicity.

(21) Systematicity should be sensitive to the organization and nature of linguistic natural kinds.

We have seen that linguistic kinds do not neatly partition the set of words in a speaker's lexicon, but are much more intricately interrelated. A successful theory of systematicity will work within the framework of these kinds, however complicated they turn out to be.

(22) Systematicity should be a contingent property of human languages.

Given the way that the systematicity of language fits into most contemporary discussions, systematicity should be an empirical property of language that a psycholinguistic model either captures or does not capture. All models capture trivialities.

(23) If only some linguistic kinds (or only some linguistic constructions) are used in a characterization of systematicity, there should be a reason why those particular kinds are used, and others are excluded.

As far as I can see, the most promising strategy for saying what systematicity is is to replace (S) with something that limits the relevant linguistic kinds and/or linguistic constructions relevant to systematicity. But in such a case we will need an argument why some kinds are relevant and others are not. It is utterly unclear to me how this might be done.

(24) It should be clear how the resulting version of systematicity is still useful and relevant for one's purposes.

A version of systematicity that meets desiderata (21)–(23) may look quite different from the unsuccessful but intuitive versions that we have reviewed. In such a case, it may be necessary to show how the successful version still accomplishes the work that one wants done from a theory of systematicity.

The desiderata in (21)–(24) form a challenge to philosophers of mind and language (and cognitive psychologists, computer scientists, and so forth) who appeal to the systematicity of language. The challenge is to construct a theory of systematicity that both respects these desiderata and yet at the same time also expresses the combinatorial completeness that is essential to systematicity. It is partly an empirical matter whether or not this can be done, so I cannot offer a definitive proof that there is no viable notion of systematicity. But when we see the kinds of troubles systematicity has, we should also ask whether the alleged systematicity of language is important or interesting enough to bother trying to theorize about. After all, whatever truth there is to the claim that language is systematic is due to other properties of the language. In particular, whatever systematicity language has is completely exhausted and explained by such things as the language's recursive combinatorial syntax (and morphology, semantics, and phonology) over linguistic constituents and the mechanisms by which we acquire new expressions by organizing them into some of a limited number of linguistic natural kinds.[27] It strikes me as plausible that

---

[27] I suspect that there is *some* truth in what Fodor and Pylyshyn say when they write that "Just as you don't find linguistic capacities that consist of the ability to understand sixty-seven unrelated sentences, so too you don't find cognitive capacities that consist of the ability to think seventy-four unrelated thoughts" (p. 40). But Fodor and Pylyshyn ignore the fact that many of the systematic-looking patterns of natural language are often cases where sentences of apparently similar form have radically dissimilar underlying structures. For instance, we might appear to have systematic variants in sentences like (i) and (ii):

(i) John promised Mary to clean the barn.
(ii) John persuaded Mary to clean the barn.

However, a closer look suggests that (i) has little in common with (ii) In fact, (i) is a better candidate for being a systematic variant of *John gave Mary a ring*; cf Richard K. Larson, "*Promise* and the Theory of Control," *Linguistic Inquiry*, XXII, 1 (1991): 103–39. Indeed, according to much current syntactic theory, (i) but not (ii) is due directly studying these features of language—how the grammar and the lexicon work—will in the long run be far more profitable than chasing after systematicity. Recursive syntax (and semantics, and so forth) and a highly structured lexicon are well-known and well-studied components of human language, whereas systematicity is at best a by-product of the former, and at worst little more than a will-o'-the-wisp.

V. ON THE SYSTEMATICITY OF THOUGHT

I turn now to a few remarks on the systematicity of thought. The notion of "a thought" is used in many different ways in the literature. I use it roughly as a catchall term for mental representations and the like. (Although this is how I use the term, my arguments will hold mutatis mutandis against many other views of thought.) Just as with language, thought is frequently assumed to be systematic.[28] Moreover, the systematicity of thought is often considered to be a fairly unproblematic datum. Peacocke, for instance, writes approvingly that "the recombinability of concepts to form new thoughts has been largely unquestioned in the published literature" (*op. cit.*, p. 42). He also announces that "This recombinability is about as general a phenomenon as one can hope to find in the realm of conceptual content" (*op. cit.*, p. 41). But the most serious problems for the systematicity of language carry over straightforwardly to thought. Thus, the first problem for the systematicity of thought is to say what it is. For instance, to maintain that the systematicity of thought is something along the lines of (S), one will need to say why (S) is the right form for thought, instead of (S′) or the form of systematicity for language, whatever that is. Second, one will need to specify the relevant mental categories, showing that each mental primitive belongs to the right mental category or categories. Finally, thought is unlike natural language in that the latter is realized as phonological (or orthographic) strings. Thought has no such overt realization, so it is even harder to determine the "parts" of a thought that fall into the various mental categories. Even if a language of thought-type mental architecture is correct, we cannot simply assume we have access to the conceptual constituents of our thoughts. After all, many of the debates in linguistics show that even trained linguists do not have that kind of grasp of all the constituents of their language. And as the history of psychology shows

---

to the same mechanisms responsible for passive constructions such as *John was kissed by Mary*. (I should note that there is only *some* truth in what Fodor and Pylyshyn say. If a being only had a small vocabulary of words available to it, its storehouse of basic sentences might indeed be quite small and unsystematic-looking.)

[28] For example, Fodor and Pylyshyn; Fodor and McLaughlin; Evans; Smolensky.

over and over, our intuitions about the machinery of our minds are very often deeply mistaken.

In short, before the ball even gets rolling there are some formidable problems with developing a clear and justified statement of the systematicity of thought. But without such an account there is nothing to discuss. I therefore assume for the time being that there are some acceptable characterizations of the components of thought, mental categories, and the systematicity of thought. In this section, I want to ask why one might think that thought is systematic. I will begin by offering a couple of reasons for doubting that thought is systematic. Since so little is known about the structure of thought, these reasons will by necessity be quite general. I will then turn to the reasons typically offered in support of the systematicity of thought. I will argue that neither reason provides any support to the systematicity of thought.[29]

There are at least two general reasons for doubting that a given empirical system is systematic. First of all, whatever precisely systematicity is, it appears to be rather rare in naturally occurring systems. For instance, the fact that one species of a genus can interact in some specified way with a feature of the environment does not imply that all species of that genus can do so too. I should mention, though, that not everyone agrees that systematicities are rare. Cummins claims that systematic systems appear to "abound in human psychology" (*op. cit.*, p. 604). Unfortunately, he gives no clear examples of this. His only putative example comes when he writes, "Consider, for example, the perception of objects in space: anyone who can see (imagine) a scene involving objects o_1 and o_2 can see (imagine) a scene in which their locations are switched" (*op. cit.*). But without specifying the psychologically relevant kinds of objects and locations, this claim is not obviously true. In this context, it seems natural to interpret "location" as location relative to other objects and background features. But although I can see (imagine) a small black box inside of a large glass sphere, I do not think I can see (imagine) a large glass sphere inside a small black box. So without a clearer understanding of what systematicity is supposed to amount to in this case, it is not at all obvious that we have found a genuinely systematic system in human psychology. In fact, the only truly systematic systems of any complexity that I know of are the languages of logic and mathematics. But these formal systems did not arise naturally; they were *designed* to have systematic properties *precisely* in order to provide a more accurate medium for reasoning than natural language (and other systems of human psychology relevant to mathematically precise cognition). But the fact that these artificial systems are systematic is no evidence whatsoever for the systematicity of a naturally occurring system like thought.

---

[29] At one point, Peacocke entertains the possibility that as a matter of stipulation, nothing would count as a thought (or an exercise of concepts) unless it exhibited the behavior characterized by systematicity (p. 48). Such maneuvers do not eliminate the empirical problem, but only shift it over. In such a case, the relevant question would shift from "Is thought systematic?" to "Do human beings have thoughts?"

The second reason to demand evidence that a given system is systematic is that systematicity is an unusually strong mathematical property. For example, the views exemplified by (S) and other standard characterizations of systematicity have the following structure.

> (25) Where C is some collection of sets of objects $\{C_1,\ldots, C_n\ldots\}$, and F is a set of functions, a collection L is systematic with respect to C and F if and only if for any $f \in F$, any k-tuple $c_1,\ldots, c_k$ of elements from $C_1 \times \ldots \times C_k$ (with each $C_j \in C$), and any $x \in C_j$ $(1 \leq j \leq k)$, $f(c_1\ldots c_k) \in L$ if and only if $f(c_1\ldots c_{j-1}\ x\ c_{j+1}\ldots c_k) \in L$.

The abstract form of systematicity could be made even more general, but (25) will suffice for now. Language was thought to be systematic, because there was thought to be some categories $C_1,\ldots, C_n$ and some syntactic operations F such that if the result of applying one of these functions to a k-tuple of expressions from the categories yields a grammatical sentence (that is, an element of L), so will the result of applying the same function to any k-tuple that differs from the original at most in that it substitutes one element for another of the same category. But notice how easy it is for systems of various sorts to fail to be systematic. For instance, let L be the even numbers, and let C_1 be the prime numbers and C_2 be the composite numbers, and let f be the addition function. This system is not systematic: for example, $f(3, 9) \in L$, but $f(3, 10) \notin L$. A bit more close to home, Frege cases appear to fail of systematicity. Let $L = \{1\}$, let C_1 be the set of words {*Hesperus, Phosphorus*}, and let $C_2 =$ {*Jones believes that* ___ *is the brightest star in the sky*}, and let $f(x,y) = 1$ if and only if the substitution of x into y yields a true sentence.

The above considerations show that we cannot just assume that thought is systematic. What evidence is there for thought's systematicity? As far as I know, only two reasons have ever been offered in support of the systematicity of thought. They are (i) that thought is similar to language in this respect (and since language is systematic, so is thought), and (ii) that it is intuitively obvious that thought is systematic. Let us examine these two reasons in turn. Does the systematicity of language provide evidence for the systematicity of thought? Pretend for the moment that we can find characterizations of systematicity for language and thought which both satisfy desiderata (21)–(24). In such a case, the systematicity of language and thought will still be independent in the sense that neither entails the other. It is possible, for instance, for a being's language to have the sort of combinatorial completeness characteristic of systematicity, even though the being's cognitive capacity lacks the corresponding property. In such a case, language as a means for expressing thought would then be like a tool that had more capabilities than its owner could ever make use of. On the other hand, a being's language might fail to be systematic even though its capacity for thought is. In this case, the being would be capable of having thoughts to which it might not be able to give perfect expression. (Or perhaps it could give expression to such thoughts, but only by using roundabout circumlocutions that did not correspond in any obvious way with the structure of the thought itself.) However, even though the systematicity of language and thought are metaphysically independent from one another, the systematicity of thought is *evidentially* dependent on the systematicity of language. When it comes to giving reasons for thinking that thought is systematic, the purported systematicity of language looms very large. We see this in the large number of appeals to the (purported) systematicity of language made at the outset of discussions of the systematicity of cognitive processing in general (compare the quotations at the beginning of this article). This fact has been noted before. Cummins, for instance, characterizes our intuitions about the systematicity of thought as follows: "The intuitions in question are claims like the following:

> [(26)] Anyone who can think that *John loves Mary* can think that *Mary loves John.*

The systematicity of thought that this is supposed to illustrate is clearly derived from the systematicity of the language—in this case, from the fact that 'John loves Mary' is a permutation of 'Mary loves John'" (*op. cit.*, p. 596). The argument here is simple: (i) language is systematic; (ii) if language is systematic, then so is thought; hence, (iii) thought is systematic. We have seen that (i) is deeply problematic. We have also seen reasons for doubting (ii), and Cummins supplies some further arguments against it. So if, as Fodor says, "the only thing in the world that is as systematic as thought is language" (*op. cit.*, p. 26), the systematicity of thought is in serious trouble as well. Since natural language is the main empirical model of systematicity to which thought is likened, the systematicity of thought is not well supported.

(Incidentally, we have surreptitiously raised another worry about premise (ii) of the above argument. It could turn out, for example, that the appropriate formulations of systematicity for language and thought are radically different. Perhaps the systematicity of thought is adequately given by (S), but the systematicity of language is given by some vastly complex version of (S′). In such a case, it is hard to see how the systematicity of one of these systems would provide any evidence whatsoever for the other.)

What about the claim that it is intuitively obvious that thought is systematic? For starters, it is hard to determine how much our intuitions about systematicity should count, since our intuitions about cognitive structure are often way off base (consider, for example, the large literature on the difference between our inferential practices and our reports of how we make various inferences). Nevertheless, I want to press the point a little, because I think it is all too often simply assumed that thought is systematic, and such a claim requires support.

Language provides a strong reason to doubt the ability of introspective evidence to support claims of systematicity. After all, many people have found it intuitively obvious that language is systematic, and we have seen that our intuitions are not to be trusted here. Furthermore, we have seen that one of the reasons that simple versions of systematicity like (S) fail is that two words can fail to behave alike because of seemingly irrelevant syntactic or semantic or phonological reasons. If Peter helped Sue carry the piano upstairs, and you wanted to explain who Sue's partner was in this joint effort, you could not use (27a), even though it simply serves to create a kind of focus on Peter from the sentence in (27b):

(27) a. *Peter is the guy who Sue and carried the piano upstairs together.
b. Sue and Peter carried the piano upstairs together.
c. Peter is the guy who Sue carried the piano upstairs with.

To make yourself clear, you might have to find some sort of circumlocution, like (27c). Thus, although you can get your idea across, the language does not allow you to do it in the simplest and most straightforward (and systematic?) way. A similar point may well hold for thought. Just as sentences are individuated by more than just their semantic content, so too thoughts (qua mental representations) may be individuated by more than just their semantic content. Moreover, just as natural language is structured around semantic categories that were unknown before the development of a scientific linguistics, so too a psychologically accurate taxonomy of semantic mental categories may be vastly different from what we pretheoretically tend to think. We cannot, for example, defend the systematicity of thought by simply *assuming* that the structure of our thoughts is very much like, for example, the structure of first-order logic.[30] Proponents of a language of thought are obviously committed to the possibility that our mental representations may have many unexpected structural properties.[31] If there is a language of thought in any interesting sense, then it may well share some similarities with every other naturally occurring language-like system. Like every other nonartificial language, the structural properties of the language of thought may simply prevent us from having certain thoughts, even when the content of that thought can be expressed by or at least closely approximated by other thoughts. If our language of thought is like English, for instance, then we may not be able to form the thought corresponding to (27a), even though we can form one with a similar content, in the form (27c). Of course, nobody knows what properties serve to individuate thoughts. Moreover, if there are structurally relevant constraints on the individuation of thoughts, it is anybody's guess what those constraints are. Indeed, it is anybody's guess what mental representations will look like as we get clearer about the relevant phenomena and develop the tools to theorize about them. The point, though, is that we cannot simply assume that the properties relevant to the systematicity of thought will be simpler than the properties that are relevant to the systematicity of language. To do so would be to adopt a very strong empirical assumption without empirical support, which is unacceptable.

VI. SYSTEMATICITY AND THE CONNECTIONISM/CLASSICISM DEBATE

The primary goal of this article has been to consider critically the notion of systematicity as such, particularly as it pertains to language. However, a natural question arises concerning the consequences of our study for the debates in which systematicity plays a part. Unfortunately, for any one of the debates a full answer to this question would be another article in itself. So I will end by merely sketching how the present view of systematicity fits into one discussion, the debate between connectionists and classicists.

The previous discussion exposes a need to clarify the role of systematicity in the connectionism/classicism debate. In addition to the questions about the form of systematicity—contrast (S) with (S′), for instance—there are also questions about what natural kinds of constituents (linguistic or mental) there are. We have seen that (given the expressions of the language) if a theory makes available enough kinds of the right sort, then systematicity is trivially true. Reflection on this fact suggests that it is not fully clear what phenomenon connectionist devices are un*able* (in some robust sense of this term) to explain without implementing a classical architecture. After all, when a connectionist system—or any kind of system, for that matter—learns language, it has to do several things, one of which is to figure out what the right linguistic (or mental) natural kinds of expressions there are, and how these kinds are organized (unless such information is "innately" given). Our discussion of systematicity has shown, among other things, that two languages could share all the same words, and even a great many of the same sentences, and yet differ primarily because they organized the words into distinct natural kinds. In fact, there are no substantial logical restrictions whatsoever on what kinds a given set of words could be organized into. But this raises the following question: Why should not every connectionist device—and indeed every language-learning device whatsoever—be thought of as having learned a fully systematic language, but as having only induced an empirically incorrect organization of natural kinds of expressions? That is, for *any* set of grammaticality judgments that a language-learning system might make, there exists an organization of kinds of words under which the language is systematic. To see this point, assume for the moment that we have adopted some form of systematicity along the lines of (S), and that we have a connectionist device *C* which is being trained to learn English. Now suppose that *C* returns the following judgments:

---

[30] Some philosophers (for example, Evans in the quote at the beginning of this article) have simply *assumed* that the structure of thought has no interesting or restrictive structure. They assume, for example, that the structure of *Mary loves John* has simply the form of a two-place relation: *Lmj*. Similarly, they might argue, the mental correlate of (27a) may be realized simply as the structure: Peter is (the x) (Sue and x carried the piano upstairs together). But adopting such a position requires making some extremely strong empirical assumptions about the structure of thought. I am unaware of any empirical evidence that supports these assumptions. I am indebted to two anonymous reviewers here.

[31] Fodor, *The Language of Thought* (Cambridge: Harvard, 1975).

(28) a. John shot Sue.
b. Sue shot John.
c. John loves Sue.
d. *Sue loves John.
e. Shot loves John.
f. *John Sue loves.

In such a case, *C* has not learned English. It does not follow, though, that *C* has not learned a systematic language. We can interpret the discrepancy between *C*'s language and English solely in terms of a difference between what linguistic natural kinds serve to structure the language and to define systematicity for that language. Thus, we can consider *C*'s language to be systematic, but the linguistic natural kinds of the language differ from those of English. (In other words, although the FGC strategy does not work for particular known natural languages like English, it does work when the nature of the learned language is unknown.) In addition to learning a systematic language, it also appears that *C* provides a kind of implementational explanation of the systematicity of the language it has learned.[22] *C* shows how to organize words into the right kinds (that is, the right kinds for that language, not for English), and thus accounts for the systematicity of that language. This interpretation of *C*'s behavior and of the language it has learned suggests that *C* has only gotten the empirical details wrong. It learned a fully systematic language, but did not get the right linguistic kinds for English. But it is not a criticism of a research program like connectionism to say only that its models have so far not gotten all the empirical details right. Getting empirical details wrong is part of the normal course of the development of any empirical research program, and indeed is to be expected in one as new as connectionism. So connectionists do not appear to be challenged to explain systematicity, but to explain how English works. But if the challenge is just to show how English could be learned, then connectionists have only been challenged to do the very best cognitive science they can. And that is not a refutation of anything.

Have I just vindicated connectionism? Absolutely not. In fact, I have not altered the evidential status of connectionism at all. I have simply shown that whatever problems connectionist models of language and cognition may face, systematicity is not one of them. There may be other ones, such as explaining how certain abilities could be innate or acquired very early and robustly, and how organization of linguistic natural kinds develops in children and so forth, but these are distinct from systematicity. At the end of section IV, I suggested that rather than trying to fix up all the problems with systematicity, we might do better in attending to the properties of natural language that give it whatever appearances of systematicity it has. These properties include its having constituents over which there is a recursive syntax, and the way in which new words enter the lexicon. Similarly, I suspect that when studying empirical models of linguistic and cognitive abilities, we might do better to ignore systematicity and study these other properties of language directly. If connectionist devices have problems modeling linguistic abilities, these problems are more likely to be exposed and understood by looking to the recursive structure of language and the way new words enter into the lexicon. After all, these are the mechanisms underlying systematicity. Any problems a theory has with systematicity will still be uncovered by focusing on these more fundamental properties. Moreover, a deeper and more principled understanding of the empirical facts will surely be useful for evaluating a given type of cognitive model. For example, many of the apparent failures of systematicity that we have looked at in this article have been cases where individual words turn out to have slightly different features that have to do with the lexicon and how words are assigned to various linguistic categories, and not with the overtly grammatical properties of language. If so, then although the property of systematicity is frequently compared to productivity—which is just the compositional property of language—in fact, it sometimes has little to do with this recursive feature. That is, many common instances of systematicity appear to have a lot to do with the lexicon and little to do with recursive syntax. At the same time, other instances of systematicity have little to do with the lexicon and much to do with syntax; further instances are an admixture of several influences. The fact that systematicity and its failures is such a mixed bag would not be at all surprising if, as I have suggested, systematicity is not a real property of natural languages, in the way that a recursive syntax is.

---

[22] I ignore many interesting further details here, such as whether *C*'s inducing the particular language it does is sufficiently robust over psychologically plausible variants of the training environment.

VII. CONCLUSION

The dominant theme of this paper has been that a theory of natural kinds of constituents is crucial to the notion of systematicity. Under a typical theory of systematicity, using standard grammatical categories, we saw that natural language is not systematic. An attempt to repair this problem by using very fine-grained grammatical categories exposes a deeper problem, namely that typical theories of systematicity make an empirically false assumption about the organization of linguistic natural kinds. The linguistic natural kinds that we have to work with form a complex overlapping structure rather than a partition. Given this system of linguistic natural kinds, it becomes unclear what it would be for natural language to be systematic. So before we can formulate an acceptable theory of systematicity, we need to answer the question "What linguistic natural kinds are there, and how are they relevant to systematicity?" The problems for the systematicity of language extend to the claims about the systematicity of thought as well, primarily because little or no substantial evidence for the systematicity of thought is offered besides thought's putative similarity to language and our intuitive feelings that thought is systematic. Furthermore, given systematicity's reliance on a theory of linguistic (or mental) natural kinds, it is unclear that connectionist devices really have a problem explaining systematicity. In fact, the explanation of systematicity appears to be nearly trivial, which suggests that the real problems lie elsewhere.

Despite all these difficulties with systematicity, it is still possible to end on a positive note. We have seen that there is a complex and challenging issue that directly concerns the nature of human language and the mind as well as the nature of systematicity, if there is such a thing. This issue is the question of what linguistic and mental natural kinds there are, and how they are organized. My guess is that a clear account of just how our expressions are organized into kinds will go a long ways towards accounting for the truth about systematicity.

KENT JOHNSON

University of California/Irvine

WHAT CAN KANT TEACH US ABOUT EMOTIONS?*

In this article I analyze Kant's theory of emotions and what he can teach us about them. I will challenge two models that have been used to explain his theory. I begin with John Sabini and Maury Silver's position,[1] according to which emotions for Kant follow the pain model: they are precognitive and involuntary phenomena. Next, I will analyze Maria Baron's objection,[2] according to which Kant held that we are responsible for our emotions. I will argue that both interpretations are misleading. First, I shall show that there is not a unique model for emotions in Kant. In his work, there is a *continuum* from uncontrollable emotions, like anger, to those that are capable of being cultivated and rationally controlled. The voluntariness and involuntariness of emotions as well as their capacity for being cultivated depends on their relation to the passive, reactive, or active self. Second, I argue that Kant's account of emotion includes both physiological aspects and cognitive contents, mainly evaluative beliefs. However, the variety of emotions presents us with different proportions of these two elements. I conclude that Kantian moral theory contributes an outstanding theory of emotions to contemporary debates, one which acknowledges physiological as well as cognitive aspects, without forgetting their diversity.

I. THE PAIN MODEL

It is a difficult task to support my thesis that Kant can teach us something about emotions and bring important contributions to the contemporary debate about this issue. Kant has long been seen as the

---

* An earlier version of this article was written during my stay in 1999 as a visiting scholar at the University of Pennsylvania, supported by a CAPES/Brazil fellowship. I would like to thank Paul Guyer for suggestions and comments on the earlier version, and the *Kant Group* of the Department of Philosophy of the University of Pennsylvania, especially Julian Wuerth, Cynthia Schossberger, Lucas Thorpe, and James Trainor. I would also like to thank Gary Hatfield and Robert Louden who helped me to improve this article and Zeljko Loparic for the incentive to keep working on emotions. I would like to thank Claudia Drucker and Luiz Henrique Dutra for a revision and discussion of the main issues of this final version, and Sofia Gollnick for help with the English. Finally, I would like to thank the University of Santa Catarina/Brazil, especially the Vice Provost for Research, Álvaro Prata.

[1] "Emotions, Responsibility, and Character," in Ferdinand Schoeman, ed., *Responsibility, Character, and the Emotions: New Essays in Moral Psychology* (New York: Cambridge, 1987), pp. 165–75.

[2] *Kantian Ethics (Almost without Apology)* (Ithaca, NY. Cornell, 1995).

0022-362X/04/0103/140–58

philosopher who denies any important role to emotions. To those who have read just the first section of the *Groundwork*, it seems to be a clear-cut case that emotions do not possess intrinsic moral value. For instance, Kant writes that, while the sympathetic philanthropist performs an action without moral worth, the absence of sympathy in the heart of the insensible one makes his action morally worthy.[3] This fact presumably inspired Sabini and Silver to claim in their article "Emotions, Responsibility, and Character," that "a Kantian chapter on emotion and responsibility is easy to write and quick to read. The domain of the moral is the domain of the will expressed in action: it is the domain of that for which we are responsible. Emotions are beyond the will, and for this reason have no intrinsic moral value" (*op. cit.*, p. 165).

According to these authors, emotions have no moral value because they follow the pain model.[4] Just like pain, which is a fact about us regardless of values or other aspects of our character, emotions, even the most complex ones, do not fall under the command of reason. Kant, just like more recent psychologists and physiologists, is said to embrace a model in which emotions are precognitive,[5] mere perceptions of unspecified bodily states[6] or undifferentiated states of the sympathetic nervous system. Emotions, like pain, would be nothing more than the stimulation of nerves, disconnected from values, character, or reason. A supporter of the pain-like conception of emotions is Robert Zajonc. In his article, "Feeling and Thinking, Preferences Need No Inference," he shows that there is a direct pathway from perceptual system to emotional responses (*op. cit.*). Accordingly, emotions are not connected with reasoning and values. According to Sabini and Silver, Zajonc's model is the same as Kant's. Both place the emotions completely out of the moral domain. Hence, we could be held responsible for the expression of our mental states, but not for having them. Besides this, in Sabini and Silver's view, certain emotional states are like the pain of withdrawal for a drug addict, in which any responsibility for action is attenuated by the intensity of the stimuli.[7]

---

[3] In the *Groundwork*, Kant writes: "Suppose that the mind of this friend of man is clouded with sorrow of his own which extinguished all sympathy with the fate of others, but he still had the power to help those in distress, though no longer stirred by the need of others because sufficiently occupied with his own and suppose that, when no longer moved by any inclination, he tears himself out of this deadly insensibility and does the action without any inclination, simply from duty, then for the first time his action has its genuine moral worth" (G 4:398). For citations of Kant's work, I will use the following abbreviations: G for *Groundwork* (1785), DV for *Doctrine of Virtue* (1797), Ant for *Anthropology from a Pragmatic Point of View* (1798), AntC for the Collins transcriptions of the *Lectures on Anthropology* (1772/73), AntPa for the Parow transcriptions of the Anthropology classes (1772/73), Rek for *On the Philosopher's Medicine of the Body*. The numbers refer to volume and page of the Academy Edition, *Kants gesammelte Schriften* (Berlin: de Gruyter, 1900).

[4] "Emotions are brute forces unconnected with higher mental functions. Pain is the obvious model. Pain is a brute force; it is beyond the will; it is, or at least typically is, independent of reason"—p. 166.

[5] Robert Zajonc, "Feeling and Thinking: Preferences Need No Inferences," *American Psychologist*, XXXV (1980): 151–75.

[6] This is the position of William James—see "What Is an Emotion?" *Mind*, IX (1884): 188–205.

Baron, in her book *Kantian Ethics (Almost without Apology)*, criticizes the Sabini and Silver position, objecting that the pain model is not the one Kant uses for explaining emotion:

> This [Kantian] model, with its very strong notion of agency, is diametrically opposed to the picture drawn by Sabini and Silver, according to which emotions function, on Kant's view, much as pain does. It is a serious mistake to think that Kant's psychology even approximately fits the model of pain and that feelings such as sympathy 'move us to act without our rational assent or assessment'.... Similarly, it would be more plausible to criticize Kant for attributing to us too much responsibility for our feelings and emotions than to attribute to him the position that we are not responsible for them (*op. cit.*, p. 197).

Baron's position is radically distinct from Sabini and Silver's: we are not passive regarding our feelings; we are responsible for the way we feel. Even if the *Critique of Practical Reason* and the *Groundwork* could justify a negative conception of emotions, as phenomena incapable of being controlled by reason, Baron claims that this conception requires revision when later texts are taken into consideration, such as *Religion within the Boundaries of Mere Reason, The Metaphysics of Morals* and *Anthropology from a Pragmatic Point of View.*

I shall try to show that neither model can fully explain Kant's view of emotions. Baron has a point in saying that we cannot explain some emotions through the pain model, like sympathy and gratitude, which can be cultivated. Kant, however, does describe some affects as involuntary and passive, such as fright in battles, which overcome the agent regardless of his will, causing disagreeable physical effects. In this case, Baron's claim that we are not passive regarding our emotions does not apply to Kantian morality. I will argue that Kant does have a robust conception of agency, in spite of our being passive regarding some emotions. If emotions are not in our power, however, this does not imply that they lack cognitive elements. People can feel anger without wanting it, but they in fact feel anger because they think something unfair or harmful has been done. Baron is right that there are some feelings like sympathy that are not beyond rational control; however, they cannot be a model for all feelings.

---

[7] More recently, Jon Elster has developed the idea of a similarity between strong emotions and addiction—see *Strong Feelings: Emotion, Addiction, and Human Behavior* (Cambridge: MIT, 1999).

My claim is that Kant presents us with a very colorful, wide range of emotions, which cannot be captured in one model type, be it a pain or a sympathy model. This diversity does not allow a simple answer concerning their voluntary or involuntary nature, or concerning the influence of physiological factors. Kant makes room for this complex picture when he connects them to our passive, active, or reactive self, as well as to superior and inferior faculties. My strategy will be, first, to present the taxonomy of emotions, and then place them in what I call a map of the self. After this first presentation of the Kantian picture, I will address some philosophical issues and relate Kant's theory to contemporary discussion on emotions.

II. EMOTIONS IN THE MAP OF THE SELF

First of all, Kant does not use the term emotion. What we preanalytically call emotion refers to at least three different kinds of phenomena: affects,[8] moral feelings, and passions. These inclinations[9] can be mainly related to two faculties: the faculty of feeling pleasure and displeasure and the faculty of desire. Feelings of pleasure or displeasure caused by an object can be sensible or intellectual. The former are caused by sensation or imagination; the latter are triggered by a concept or idea (Ant 7:230). Pleasure and displeasure given by sensibility alone are feelings of gratification and pain. Sensible pleasure and displeasure admit also two other kinds of inclination, caused by imagination: sympathy and affects. The difference between them is that sympathy can be cultivated, while we are passive with respect to affects.[10]

---

[8] I use affects and passions as translations for *Affekten* and *Leidenschaften*. I use moral feelings as referring to sympathy, moral feeling itself (*Moralische Gefühl*), and love of benevolence. I claimed in a previous paper that in later works, like *Anthropology* and the *Doctrine of Virtue*, Kant has taken sympathy as a moral feeling, although in his previous *Lectures on Anthropology*, as well as in the *Groundwork*, sympathy seems to be taken as an affect—see Borges, "Sympathy in Kant's Moral Theory," in Volker Gerhardt, Rolf-Peter Horstmann, and Ralph Schumacher, eds., *Kant und die Berliner Aufklärung*, Akten des IX. Internationaler Kant-Kongresses (Berlin: de Gruyter, 2001), band III, pp. 152–58.

[9] I use inclination here in a wide sense, referring to a wide range of empirical impulses, including instincts, affects, appetites. Inclination can also have a specific meaning in Kant's thought, as the third degree of the faculty of desire For a more detailed explanation about inclinations, see Allen Wood, *Kant's Ethical Thought* (New York: Cambridge, 1999), and Robert Louden, *Kant's Impure Ethics: From Rational Beings to Human Beings* (New York: Oxford, 2000).

Affects are feelings of pleasure or displeasure that hinder the reflection through which inclinations were to be submitted to rational maxims; they are sudden and rash, making reflection impossible (DV 6:408), such as water that breaks through a dam or a stroke of apoplexy (Ant 7:252). They can lead the agent to moral blindness, since they hinder deliberation, with the consolation that this tempest easily goes away and calms itself, allowing the subject to go back to a state where reflection is possible again. Kant cites the example of someone who marries out of love and is blind to the flaws in the character of her beloved, but regains her vision a week after marriage (Ant 7:253). The Kantian paradigmatic example for affect is anger, a tempestuous feeling by nature, and fickle like love. Sympathy, although it is a sensible feeling,[11] can be trained in order to help the accomplishment of moral actions, when respect for the moral law is not a sufficient incentive.

Pleasures and displeasures are not all sensible, however; we also have the intellectual ones, caused by an idea or concept. These include feelings that make the mind receptive to the concept of duty, such as moral feeling, which is defined as "a susceptibility to feel pleasure or displeasure from the consciousness that our actions are consistent or contrary to the laws of duty" (DV 6:399).[12] Love for human beings belongs also to the receptivity of the mind to the concept of duty. Intellectual pleasures include love of benevolence, but not love of satisfaction. The former but not the latter could be ordered as a duty, since it would be contradictory to have a duty to feel pleasure. Love of benevolence admits of something like Aristotelian cultivation, for it is a disposition that can be awakened by habit.[13]

---

[10] In *Anthropology from a Pragmatic Point of View*, Kant refers to this kind of sympathy as sensitivity, a capacity to feel states of pleasure or displeasure with the happiness or misery of the others. Kant writes: "Sensitivity is not opposed to equanimity because it is a faculty and a power which either permits the states of pleasure or displeasure, or even keeps them from being felt. Sensitivity is accompanied by a choice. Sentimentality, on the other hand, is a weakness because of its interest in the condition of the others who could play the sentimentalist at will, and even affect that person against his will" (Ant 7:236)

[11] See DV 6:456.

[12] Love of human beings also belongs to what Kant calls "[c]oncepts of what is presupposed on the part of feeling by the mind's receptivity to concepts of duty as such." However, Kant denies that this kind of practical love is, properly speaking, an emotion, since we cannot have the duty to love someone, if love is a matter of feeling. One can only cultivate this practice of benevolence, but to call it love is "very inappropriate" (DV 6:401).

[13] See DV 6:402.

Kant's realm of inclinations also includes passion, which is related to the faculty of desire and refers to a strong desire for something. Passions exhibit a contradictory nature. On the one hand, Kant says that they hinder the control of reason to compare, at a particular moment, a specific inclination against the sum of all inclinations (Ant 7:265). On the other hand, they admit some rational deliberation about the means to obtain what the agent desires. One good case is given in the *Anthropology*, where Kant compares the inability of a man who feels the affect of love to seduce someone, to the skill of one who is taken by the passion of love. The first will not be successful, while the second can easily trap the helpless victim (Ant 6:265). The difference is that one is immersed in a full agitation of the mind, while the other, in cold blood, plots the way to obtain it. This case depicts the difference Kant makes between the meaning of affects and passion, a difference that is also illustrated in outstanding literary works, such as *Dangerous Liaisons*.[14] While affects are outbursts of feelings, which cannot coexist even with a prudential rationality, passions show their cunning. The same difference can be seen when one compares the affect of hate with the passion of anger: while the latter is fickle, the former is a permanent disposition, which inclines the agent to plot a cold-blooded vengeance.

In order to improve our understanding of Kant's model or models for emotions, it is helpful to locate them in relation to a Kantian sketch of the self.[15] Kant claims there are three instances of soul[16] in a generic sense, whose reference in each case is the self. The self can be observed in a triple perspective: as passive, reactive, or purely active. This tripartite division of the self is connected with the superior and inferior faculties. The inferior faculty corresponds to the passive part; the superior faculty, as active, corresponds to the perspective of both a purely active self and a reactive one.

The superior/inferior distinction applies to all three faculties: the cognitive faculty, the faculty of desire, and the feeling of pleasure and displeasure. As regards the faculty of feeling, we have pleasure or displeasure of sensibility alone through the inferior faculty, while the pleasures, given by imagination and understanding, belong to the superior faculty. This difference can be illustrated when one contrasts pain, which belongs to the inferior faculty of feeling, to the feelings of pleasure and displeasure in the superior faculty, such as moral feeling. Physical pain seems not to allow a rational control, because it is displeasure of the animal or passive soul. Pain is defined as "displeasure of sensation," and is explained by the effect produced in the mind by the sensation of one's physical condition. Totally opposed to physical pain, we have the pleasure or displeasure due to a concept, such as the moral feeling of *The Metaphysics of Morals*. This feeling is an effect of the concept of duty: the moral feeling "is the susceptibility to feel pleasure or displeasure merely from being aware that our actions are consistent with or contrary to the law of duty" (DV 6:399). Hence, there is a big contrast between pain and moral sentiments, even though they could both be classified as feelings: the first one is related to the inferior faculty of feeling, connected with sensibility; the second one is related to the superior one, connected with reason. While physical pain is an involuntary, precognitive feeling, moral feeling includes the concept of a right action. We cannot decide whether we will feel pain or not, while we can decide whether we will feel a moral pleasure, because this requires just that we act according to the moral law.

---

[14] See Choderlos de Laclos, *Les Liaisons Dangereuses* (Paris: Gallimard, 1972).

[15] I am indebted to Julian Wuerth and his outstanding work for the sketch of a map of the self—see Wuerth, "Kant's Account of the Self" (Ph.D. diss., University of Pennsylvania, 2000) However, the division of emotions in the instances of the self is my own responsibility

[16] According to Collins's notes on the *Anthropology* (1772/73) we observe the soul in a triple perspective, like soul (*anima/Seele*), mind (*animus/Gemuth*), and spirit (*mens/Geist*). The passive soul is called soul/anima (*Seele*); the reactive soul is mind/animus (*Gemuth*); while the purely active soul is spirit (*mens/Geist*). These three instances of the soul refer not to three substances, but to "three forms of feeling alive." See Collins's notes on the *Anthropology*, AntC 25:16. For a groundbreaking and insightful analysis of the notes on *Anthropology* concerning the self, see Wuerth.

The inferior faculty of pleasure is responsible for purely sensible phenomena such as pain, hunger, and thirst. The superior faculty of pleasure relates to the reactive and active self. The sensation of displeasure that we feel when we know that our actions are wrong is associated with the active self, since it is a feeling connected with a concept, the concept of duty. Between the pure passive and the pure active part of the self abide the reactive feelings, which still belong to the superior faculty of feeling, and refer to affects and sensibility.

Kant, who was a well-known hypochondriac, illustrates in his *Lectures on Anthropology* the difference between passive, reactive, and active feelings with a case of a man who suffers from gout:

> I cannot prevent the pain inflicted on my body from passing into my soul. I can only prevent that my soul reflects over this, e.g., when I have gout and think what will become of me in the future, how I will acquire my bread and this causes sadness over the state of my health, here *animus* agitates. The sickness of the mind is also what makes me miserable. Because such reflection never attaches to animals, they are never miserable. But finally the highest degree of sadness arises when my spirit abstracts from all pain and awakens in me a self-reproach, when it

> imagines to itself how I brought this illness upon myself and became unhappy through my own fault (AntPa 25:247–48).

We can clearly see, in this quotation, the difference between three levels of displeasure. The first one is purely physical, beyond control, even indirectly. Since it is purely sensible, it is independent of any cognitive content; the subject feels a certain wound as pain. This is an example of the inferior faculty of pleasure and displeasure, where the feeling is given through sensation alone.

The second level relates to the displeasure of an affect, sadness, and allows a nonphysical cause: sadness is caused by the imagination that agitates the mind when the agent worries about the future. Here we have the reactive part of the self, which relates to the superior faculty of feeling. Sadness is aroused by imagination, which agitates the mind. Animals, Kant argues, are never miserable, although they can feel the pain of a wound. However, Kant claims in another lecture, *On Philosophers' Medicine of the Body,* that animals do have imagination, although in cattle "this force [of imagination] is not directed by any choice or deliberate intention of animal." Since some animals do have imagination, Kant can conceive that a certain illness may oppress the mind of the animal when it is brought into captivity, "yet the black anxiety that afflicts the miserable human race escapes the animal, which knows nothing of worry."[17] To mention animal mind, which can be affected by sadness, *ipso facto* refutes the idea that Kantian affects fit the pain model; since even animals are supposed to have affects triggered by imagination. The very idea of being in captivity can trigger in their minds some emotions of sadness. In human beings, this faculty is supposed to be even stronger, giving a deeper force to affects. Animals may feel sadness, although their weaker imagination does not allow for deep anxiety. Their emotions, according to Kant's account, are not always involuntary, nor are they always precognitive. The idea of captivity brings them the idea of a miserable future, which triggers the affect of sadness. Thus, the pain model will not even explain the emotions Kant ascribes to animals.

Humans have another level of emotions, the one connected with moral judgments. This third level, the displeasure of spirit, the pure active part of the self, is completely absent in animals, since it depends on reason that awakens self-reproach. The moral conscience would create in the agent the feeling of displeasure because she knows that she has not acted well. Self-reproach is, then, an illustration of an intellectual displeasure, caused just by a concept. As we have seen, in the classification of the faculty of pleasure and displeasure (Ant 7:230), we have the division between sensuous and intellectual pleasure. While the former can be produced by sensation or imagination, the latter is conceived by concepts or by ideas. In the case of self-reproach, displeasure is produced by the idea of moral law.

---

[17] See Rek 15:944.

Passions, which belong to the faculty of desire, are also related to the active self, since agents form maxims according to their passions. The thirst for vengeance, evil as it may appear, is still a maxim of reason, developed out of an injustice suffered. History gives many illustrations of plotted murders based on hatred, sometimes based on ambition and lust for power. Such passions lead the mind to accept evil maxims, in order to accomplish its ends. That is why Kant claims that passions manifest certain characteristics of reason (Ant 7:270).

What should we conclude from this taxonomy of emotions?[18] Since the pain model is obviously ruled out, should we assume that Baron is right? Are we, according to Kant, responsible for our emotions? Since Kant supposes that humans can have moral feelings of displeasure according to the rightness or wrongness of our actions, is it the case that we do have control over all emotions? The answer is negative—moral feelings are but one part of the wide range of emotions. The nature of moral feeling, as something produced by the concept of moral correctness, hence controllable, arises from its be-

---

[18] This reconstruction gives us the following table, where emotions are in bold letters:

Faculty of feeling pleasure or displeasure			Faculty of desire			
Inferior faculty of feeling	Superior faculty of feeling		Propensity	Instinct	Inclinations	**Passions (hate, greed, lust of power, ambition)**
Passive self soul, *anima*, *Seele*	Reactive self mind, *animus*, *Gemuth*	Active self spirit, *mens*, *Geist*				
Pain, hunger	**Affects (anger, sadness, joy)**	**Moral feelings**				

longing to the active part of the self. Nevertheless, between the purely passive part, where the feeling of pleasure depends on the form through which the object affects sensibility, and the purely active part, in which the satisfaction should relate directly to a concept, we have a variety of intermediary phenomena, such as sympathy and affects.

Sympathy and affects belong to the reactive realm of the soul. They are phenomena that are different from pain, in that they require other faculties besides sensation, and they are different from moral feelings, in that they are not provoked by a concept. That is why neither the pain model nor Baron's model can explain them.

The affects are typically reactive agitations, or reactions to something that has affected our mind. Kant's view of these feelings, which include anger, joy, sadness, and fear, is usually negative: affect acts like a surprise through sensation, suspending the composure of mind. Therefore, emotions are precipitate and also have the unwanted property of growing to a degree of feeling that makes reflection impossible (Ant 7:252).

Being like a surprise, this kind of feeling cannot be directly controlled by the will. We can decide whether or not to perform a correct action, and hence whether we will feel moral pleasure or displeasure resulting from such action. But we cannot decide whether it is appropriate to feel anger in a particular situation. Sometimes, we feel anger, even if we do not desire to feel it; other times we can think it would be fair to feel it, although we are insensible in the situation. Kant illustrates this possibility in the *Anthropology*, mentioning Socrates, who was in doubt whether it would be good to be angry sometimes, since it is paradoxical to have emotion so much under control that one can cold-bloodedly deliberate whether or not one ought to be angry (Ant 7:253). We cannot deliberate about feeling angry, although anger does not lack cognitive content. The affects have fewer cognitive elements than moral feelings; however, they have more cognitive elements than pain or other mere physical feelings. Anger involves the perception that something unfair was done, which hurts the agent or is against her conception of justice. We do not feel anger the way we feel pain; we feel anger when we realize that something in a situation is offensive or unfair.

Sympathy is a sensible feeling that admits of choice. It can be cultivated, in order to give a correct response in situations where we need practical benevolence. It is a phenomenon of the reactive part of the soul, which can be changed by the active part. Sympathy, as referred to by Kant in the *Doctrine of Virtue*, fits the conception of sensitivity mentioned in the *Anthropology*. It is a natural feeling, which should be cultivated and used when the respect for moral law is not sufficient to trigger moral actions.

Feelings from the animal (or passive) part of the self are involuntary, and cannot have their sensation controlled. This is the case with pain, thirst, hunger, and so on. In the other extreme, we have the pleasure and displeasure of the active self, such as moral feelings and the feeling of respect, which are controllable through our actions, since they are outcomes of good or evil actions. In between lies affect, a phenomenon of the reactive part of the self, which is connected with the imagination.

When we mention reactive feelings, such as affects, we should consider also that some strong affects have involuntary outcomes, which involve strong physiological arousal. Kant cites two highly illustrative cases in this connection, related to anger and fright. About anger, he mentions the situation of a man who enters one's room in anger in order to say harsh words (Ant 7:252). In this case, one should make the angry man sit down, because the comfort of sitting is a relaxation, which does not conform to a great wrath. As for fright, he mentions the disagreeable digestive effects that can affect soldiers before a battle (Ant 7:257). Kant's account of the uncontrollable and even highly undesirable outcomes of these emotions agrees with contemporary findings in physiology, according to which emotions may cause effects related to the autonomic neural system.[19] These physiological components of emotions are responsible for the difficulty in controlling them,[20] since they have an inertial component, which is not easy to handle.

III. BETWEEN THE PROPOSITIONAL ATTITUDE SCHOOL AND THE FEELING THEORY

For Kant, emotions do have evaluative components, but, with the exception of passion, they are also feelings, which have physiological features. This is why Kant recommends that we should soothe the movements of a mad man, making him sit down and relax. The soothing of movements will help to calm the affect of anger itself.

---

[19] These effects include sympathetic system effects, such as acceleration of the heart, and parasympathetic system effects, such as loosening of the bowels. See Ronald de Sousa, *The Rationality of Emotions* (Cambridge: MIT, 1997), pp. 47–52; Elster, p. 36.

[20] According to Elster, some emotions, like love and anger, are characterized by a point of no return beyond which self-control is no longer possible, since we lose control before we become aware of the emotion—see Elster, pp 28, 29. I would like to suggest both that this no return point is due to inertial physiological aspects, and that Kant was aware of this when he said that the man in anger should sit down to calm himself. Once anger is triggered, the moment one notices it, it is already too late to maintain self-control

We can say that Kant is between the propositional attitude school and the feeling theory.[21] According to Paul Griffiths,[22] the former explains emotional phenomena mainly by beliefs and desires,[23] while the latter claims that emotions are characterized by a quality and intensity of sensation. Kant proposes a view of the emotions according to which they are intentional states as well as feelings. They are intentional states[24] in so far as they are about something, or directed to something in the world beyond themselves. We are not only mad, as we feel pain, but we are mad with someone or at something, we are not in love without an intentional object, we are in love with someone. As intentional states, they have a propositional content, and also bring evaluation and cognitive elements; as feelings they present physiological arousal. In this way, he can overcome the dichotomy of thought and feeling. Kant will not deny that an emotion has mental content. However, it is not only a desire for, or a judgment *about* something. Physiological processes also accompany it.

Kant would agree with the propositional attitude school, in that there is no sense in attributing sadness without the idea that something valuable was lost, or anger without the idea of something offensive done by someone. We are angry *with* a specific person or *at* a determinate situation; anger consequently has an intentional object. However, he seems to acknowledge that for some strong emotions, such as anger, a physiological arousal seems to be a necessary condition for emotion. In the case of anger, the man has to sit in order to calm himself, because anger is accompanied by a strong physiological arousal, which cannot be instantly overcome by the sole reasoning about the emotion.

---

[21] The feeling theory goes back to Descartes, Hume, and James in "What Is an Emotion?" Recently, the work of Zajonc emphasizes the noncognitive elements of the feeling theory.

[22] See *What Emotions Really Are* (Chicago: University Press, 1997), pp. 1–24.

[23] Examples of propositional attitude theory can be seen in Anthony Kenny, *Action, Emotion, and Will* (London: Routledge Kegan Paul, 1963), and Robert C. Solomon, *The Passions* (Garden City, NY: Anchor, 1977). The propositional attitude school could also be labeled in a more general way, as cognitive theory. According to de Sousa, cognitive theories maintain that emotions have informational content. In his view, Solomon's theory is a modified cognitive theory, where the judgments involved are of a special kind, they are evaluative ones—see de Sousa, p. 40.

[24] John Searle, in his last book offers a good summary of the basic structure of intentionality—see "The Basic Structure of Intentionality, Action, and Meaning," in *Rationality on Action* (Cambridge: MIT, 2001), pp. 33–60, on p. 34. I will take his definition of intentionality here: "Intentionality as philosophers use the word, refers to that aspect of mental states by which they are directed at, or about, or states of affairs in the world beyond themselves." Although Kant did not use the vocabulary of intentional states, I think he would agree with Searle that emotions are propositional states.

Will this physiological arousal also accompany emotions related to the active part of the self, such as moral feelings? The answer is yes. Although moral feelings are ultimately caused by an idea, the idea of moral correction of an action, they are still feelings of pleasure and displeasure. Although here the physiological arousal is not as strong as in anger, the idea of feelings brings with it pleasure and displeasure, which is impossible without any physiological component.

All emotions that are feelings of pleasure and displeasure have both components: evaluative contents and physiological arousal. Emotions express beliefs. In the case of the man who suffers from gout, his sadness is related to the belief that a gloomy future is waiting for him. His feeling guilty is related to the belief that he has contributed to his actual misery. The pain model does not apply to the Kantian account of emotions, because beliefs and desires are constitutive of these mental states. However, Baron's picture is not accurate either, since emotions are also composed by physiological components, which are scarcely under our control.

What kind of cognition is involved in emotions? Kant will certainly deny that emotions are mere reflexes, like pain, without any moral evaluation of a state of affairs. Some affects involve moral assessment, and they can be valuable in giving us information about the moral salience of a situation.[25] Sympathy can give us information about someone in distress. However, to ascribe cognitive value to emotions is not to say they have normative value. Emotions cannot tell us what to do; this is the task only of reason. Sympathy could inform us about the existence of someone in need, but only reason can tell if it is morally correct to help this person in that situation. Sympathy is connected with the belief that someone needs our help. Yet this emotion cannot validate for itself that it is right to help that person. A student can feel sympathy for a colleague who cannot do his own assignment, and want to do it for him. Reason will tell him this is not the right thing to do. Anger could inform us that someone did us some harm and trigger the desire for vengeance. Yet reason will tell that vengeance is not the right thing to seek. As Michael Moore correctly remarks, for Kantian morality "the emotions that generate a moral belief are irrelevant to the truth of that belief."[26] In that sense, good emotions could give rise to false as well as to true moral beliefs. This does not imply that emotions are not connected with beliefs. In the end, what Kant offers us is the more subtle idea that emotions are really connected with moral assessment, yet a moral command based solely on emotion cannot be validated without further rational considerations. Our compassion for a murderer who has had an awful childhood could let us forgive his crimes; our sympathy for a friend who suffers could lead us to not telling her a sad truth. Yet these are not morally right courses of action.

---

[25] About moral salience and emotions in Kant, see Barbara Herman, *The Practice of Moral Judgment* (Cambridge: Harvard, 1993).

[26] "The Moral Worth of Retribution," in Schoeman, ed., pp. 179–219, see p. 198

If we take into account that passions can also be put under the *label* of emotions, then here we have a different relation between emotions and morality. While affects hinder momentarily the sound use of reason, passions are always evil. They also have a different relation with physiological arousal, and seem not to exhibit them just like feelings do. Thirst for vengeance, greed, lust for power, or ambition can be accompanied by no physiological arousal, since they have characteristics of cold-blooded reason. They are more desires then feelings, having a salient intentional object and lacking physiological disturbances.

We can still say that the majority of emotions usually have both intentional object and physiological arousal, since for the most part they are feelings. However, passions lack physiological arousal and can be conceived rather as a strong desire for something, being then defined more specifically as propositional attitudes than as internal feelings. Kant shows sophistication in dividing what we usually call emotions into two kinds of events: feelings of pleasure and displeasure and passions, both of which are composed by different components.

IV. ARE EMOTIONS USELESS FOR KANTIAN MORALITY?

Emotions have cognitive and evaluative content; however, they are not trustworthy in showing us what to do. Is the cognitive content of emotions therefore useless for morality? I do not think this is Kant's answer, since he recommends, for instance, that one should visit places where people suffer in order to awake our natural sympathy. He presents a more sophisticated approach, where the content of our emotions should be scrutinized by reason.

What does the Kantian account of emotions teach us about the relations between feelings and morality? First, emotions do not constitute a radically different realm from cognition and morality. For Sabini and Silver, on the other hand, morality and emotions constitute two different realms, and for Baron they are concurrent domains. Radical difference, supported by Sabini and Silver, would make emotions irrational feelings; the total coincidence (Baron) would imply both that we can be considered responsible for what we feel, and that emotions can be easily controlled by reason. The former conception is wrong because texts such as the *Doctrine of Virtue* explicitly show how we could and should cultivate some emotions in order to accomplish benevolent actions, when respect for the moral law is not a sufficient incentive. But Baron's statement that we are responsible for the way we feel is also not true, because there are some emotions like fright and anger, that arise in ourselves without permission and do not permit easy control. Hence, the relations Kant establishes between morality and emotion are more complex than total distinction or total similarity.

There is no unique answer regarding the role of emotions in morality, since they have a multiple reference to a *continuum* ranging from reactive phenomena to active phenomena of the self. An example of the latter is moral feeling, which is generated by the consciousness of the moral value of an action. Most of what we call emotions, however, lies in the reactive part of the soul. But here also we have a *continuum*: from affects, which are the most uncontrollable, such as anger or fright, to sympathy, which can be modulated by the will in order to help the accomplishment of moral actions. At one extreme, we have anger and fright, at the other, sympathy and moral feelings.

The mistake of commentators is to consider that emotions in Kant have only one model and refer to only one kind of event, when, in fact, they refer to a multiplicity of different phenomena, which demand different explanations. These mistakes in defining a correct concept of emotion result in a misunderstanding of its role in morality. Anger can bring with it moral assessment about the unfairness of a situation; however, since it is an affect, it will make deliberation more difficult and will be a hindrance to moral action. Sympathy, although sharing with affects some features, because it is still a feeling of the reactive self, can be shaped by the will in order to help morality. Moral feeling, as a feeling of the active self, is directly connected with morality and with the moral evaluation of a situation.

Passions have a special relation with morality: they are, with no exception, evil. They are not, like affects, only agitations of the soul, which hinder reflection and moral deliberation. Moreover, some affects can give us moral information about the situation, although immoderate and difficult to control. Passions, however, are quite the opposite: they are immoral desires for something. Thirst for vengeance and power, greed and ambition present a perversion of moral goodness, since they are related to an evil deliberation of a mind that is not disturbed by affections.

V. THE ROLE OF IMAGINATION

One of the problems regarding the propositional attitude theory is that it cannot fully acknowledge the role imagination plays in the production of emotions. Kant makes room for imagination in his account. Emotions are not feelings without any cognitive content; therefore they do not belong to the realm of total involuntariness. They are feelings that can be awakened in us through imagination or reason, produced by the reactive or active self.

Imagination plays an important role in the arousal of reactive affects. Sadness, as we saw, results from the imagination of a future unwanted occurrence. It is related, like joy, to the reproduction or anticipation of events. Kant cites the homesickness of the Swiss as an illustration of the role of imagination: the Swiss, when transferred to other lands, have the feeling of homesickness, which is aroused by the recollection of places where they enjoyed the very simple pleasures of life (Ant 7:178). Here, the recollection of good moments in the past can produce this feeling. In the case of the sick man, he becomes sad when he imagines a bad future and worries how he would acquire his bread. Imagination can intensify the feeling of sadness, as the example of animals that are sad, but never miserable, shows, because they do not have the intensity of imagination human beings have.

In the text *On the Philosophers' Medicine of the Body*, Kant ascribes to imagination an even stronger force: the power of imagination can heal or produce a disease. On the one hand, the confidence sick people put in their doctors helps them through the imagination that they will be healed. On the other hand, imagination could spread diseases like epilepsy, since it is able to alter bodily movements.[27] Since affects are movements of the mind that agitate the body (Rek 15:940) and imagination can alter bodily movements, it can also interfere with affects. However, this alteration should not be considered as an effect of the direct power of the will.

Imagination can modify affects by bringing images to mind that cause agreeable or disagreeable sensations. The role of imagination is clear in the case of fear: just like sadness, which is caused by the anticipation of a bad future, fear is caused by the anticipation of danger. These feelings, even though they belong to the reactive part of the soul, have some cognitive elements that go into the evaluation of images that come before the mind. The idea that affects are blind and completely precognitive is wrong. Feelings of sadness, joy, anger, and fear depend upon an appraisal of a situation reproduced or anticipated by imagination.

---

[27] Kant writes: "That's why the hideous movements that we call convulsions, and epilepsy, can spread to other people: imagination gives them a certain contagiousness. This gives rise to a strategy, which does not belong to the doctor as such, of practicing medicine merely through the force of imagination.... Hence the confidence invalids put in the doctor adds strength to his remedies for even the most enfeebled" (Rek 15:944).

We cannot infer, however, from the fact that there is a cognitive judgment of images, that they are easily controllable, which is a mistake of other commentators who take sympathy as a model for all affects. Taking again the case of the *Lectures on Anthropology*: I could decide that a state of sadness is bad for my disease, even though, each time I think about my future, I become sad, against my will. However, if the sick man suddenly remembers that he has a million dollars in health insurance for gout, his future would suddenly appear to him as a nicer one, and these images of a good and nice future will arouse in him the feeling of relief. Moral feelings, like the pleasure or displeasure in the rightness or wrongness of our conduct, can be more related to judgment than to imagination. If the sick man were to reproach himself for his situation in the former case, now that he remembers he has health insurance, he could feel pleasure and self-satisfaction in being prudent and wise.

VI. CAN WE BE HELD RESPONSIBLE FOR OUR EMOTIONS AND ACTIONS?

Can we be held responsible for our emotions? It is true that some emotions of the reactive self, such as joy, sadness, and longing, are intimately connected with imagination and this faculty does have the power to modulate affects. However, the power of imagination cannot deny the evidence for my belief in a bad or good future. Although we can modulate some affects through imagination, we cannot decide not to have them or to have them at ease. Imagination of a future state of pleasure or displeasure should be connected with a possible state of affairs. If the sick man could not work in the future because of his gout, he cannot prevent the feeling of sadness for the bad future he imagines. Or, he can even trigger for some moments the affect of joy when he remembers his past moments, but as soon as he focuses on his future situation, this joy will be quickly turned into sadness.

If we are not responsible for our emotions, can we be responsible for our actions? Emotions are said to be an obstacle to moral deliberation, and Kant uses phrases such as *water that breaks a dam* in order to show the irrational force of these feelings. However, he also draws on a strong picture of freedom, which does not allow for pathological compulsion.[28] We have strong emotions, difficult to control, but we have means to tame them in order to act the way we want. Virtue is one of these means. We are not responsible for emotions, although we can be held responsible for actions, since the strength of emotions cannot be equated to compulsion, such as the compulsion for drugs. Recently, some philosophers have tried to establish this kind of parallel between strong emotions and addiction.[29] Kant would not accept this picture, because as strong as emotions can be, and as much of a problem for morality they can portray, the very idea of practical reason presupposes that agents can decide how to act. In fact, the Kantian picture here is more likely to be accepted by moral common sense, since a strong emotion can never be taken as a total excuse for a bad action. Agents can mention strong anger to explain their violent acts, yet not to forgive them. People are held responsible for wrong actions due to strong emotions, because it is presupposed that they could have acted otherwise.

VII. WHAT CAN KANT REALLY TEACH US ABOUT EMOTIONS?

In order to explain emotions, philosophers have tried, with few exceptions,[30] to build a single model for something that they supposed was a single kind of mental event. But perhaps the key to the understanding of emotions is to grasp their complexity and the different weight of rational and irrational components each one has. Although Kant did not write one specific book on emotion, I think he was very attentive to this variety of emotional events, particularly in his Anthropological works.

In this brief reconstruction of what could be a Kantian theory of emotion and its relation to the passive, reactive, and active self, I have shown that this kind of feeling presents different relations with our passive and active self. Emotions cannot be explained by the pain model, because they are not outside the rational domain, since they are also composed of evaluative beliefs and judgments. They are not involuntary, nor are we responsible for feeling them. They are phenomena of the reactive and active self, which can be modulated by imagination or by reason, but we cannot provoke or extirpate them by the sole power of the will. We can be held responsible for our acts, but not for our emotions, which is what makes Baron's model mistaken.

---

[28] See *Lectures on Ethics*, Mrongovius: "Can I really conceive of a pathological compulsion in man as well? Truly I cannot, for freedom consists in this, that he can be without compulsion in the pathological sense; nor should he be compelled in that way. Even if a man is so constrained, he can nevertheless act otherwise" (29:618).

[29] See Elster.

[30] Some of the recent exceptions are Griffiths and Elster, who seem to acknowledge that emotions are not a single type of event. Yet, from this fact, they have reached different conclusions. While Griffiths maintains that the category of emotion is not useful philosophically speaking, Elster seems to accept the philosophical use of this category.

Kant's model accepts features of both the propositional attitude model and the feeling model. From the first, it takes the idea that emotions have an intentional object, and from the feeling model, it takes the idea that emotions are accompanied by physiological arousal. Emotions involve evaluative beliefs and also physiological disturbances. The degree of physiological arousal depends on whether emotions belong to the active or reactive self. And among the later ones, there are emotions whose physiological arousals are stronger than others. This is the case with anger, if compared to the mild affect of longing or sadness.

Kant's theory is more likely to be accepted by our common moral sense, because it explains emotions as people actually feel them. It is a very sophisticated theory that confirms the complexity of its object and does not allow for unrefined explanations. And perhaps he shows his cunning in not using the word emotion, but instead, other terms such as *affects*, *sympathy*, *moral feelings*, and *passions*, in order to show that what we preanalytically call emotions refer to a wide variety of states, which call for different philosophical categories. This insight agrees with some contemporary philosophers who have recently challenged the idea that emotions form a unique class of events.

MARIA BORGES

University of Santa Catarina/Brazil

NEW BOOKS

KIVY, PETER: *Introduction to a Philosophy of Music.* New York: Oxford, 2003. ix, 283 p. Paper $21.95; *The Seventh Sense: Francis Hutcheson and Eighteenth-century British Aesthetics,* Second Edition. New York: Oxford, 2003. xiv, 402 p. Cloth $74.00, paper $24.95.

KLOSKO, GEORGE: *Jacobins and Utopians: The Political Theory of Fundamental Moral Reform.* Notre Dame: University Press, 2003. xii, 200 p. Cloth $35.00, paper $17.00.

KORNBLITH, HILARY: *Knowledge and Its Place in Nature.* New York: Oxford, 2003. x, 189 p. Cloth $29.95.

KRAMER, MATTHEW H.: *The Quality of Freedom.* New York: Oxford, 2003. ix, 481 p. Cloth $45.00.

KUNTZ, MARION LEATHERS: *The Anointment of Dionisio: Prophecy and Politics in Renaissance Italy.* University Park: Penn State UP, 2001. xviii, 446 p. Cloth $55.00.

KUSCH, MARTIN: *Knowledge by Agreement: The Programme of Communitarian Epistemology.* New York: Oxford, 2002. xiii, 303 p. Cloth $45.00.

KVANVIG, JONATHAN L.: *The Value of Knowledge and the Pursuit of Understanding.* New York: Cambridge, 2003. xvi, 216 p. Cloth $60.00.

LAHEY, STEPHEN E.: *Philosophy and Politics in the Thought of John Wyclif.* New York: Cambridge, 2003. x, 237 p. Cloth $60.00.

LAKOFF, GEORGE, and MARK JOHNSON: *Metaphors We Live By,* Updated Edition. Chicago: University Press, 2003. xiii, 276 p. Paper $14.00.

LAMBERT, KAREL: *Free Logic: Selected Essays.* New York: Cambridge, 2003. xii, 191 p. Cloth $55.00.

LANDESMAN, CHARLES: *Skepticism: The Central Issue.* Malden, MA: Blackwell, 2002. x, 210 p. Cloth $59.95.

LAWLOR, LEONARD: *Thinking through French Philosophy: The Being of the Question.* Bloomington: Indiana UP, 2003. xvi, 212 p. Cloth $59.95, paper $24.95.

LECERCLE, JEAN-JACQUES: *Deleuze and Language.* New York: Palgrave Macmillan, 2002. ix, 274 p. Cloth $68.00.

LEDBETTER, GRACE M.: *Poetics before Plato: Interpretation and Authority in Early Greek Theories of Poetry.* Princeton: University Press, 2003. viii, 128 p. Cloth $29.95.

LEE, DANIEL E.: *Navigating Right and Wrong: Ethical Decision Making in a Pluralistic Age.* Lanham, MD: Rowman and Littlefield, 2002. xv, 143 p. Cloth $55.00, paper $16.95.

LOSONSKY, MICHAEL: *Enlightenment and Action from Descartes to Kant.* New York: Cambridge, 2003. xvii, 221 p. Cloth $60.00.

LUFT, SANDRA RUDNICK: *Vico's Uncanny Humanism: Reading the "New Science" between Modern and Postmodern.* Ithaca, NY: Cornell, 2003. xviii, 213 p. Cloth $45.00.

LUKES, STEVEN: *Liberals and Cannibals: The Implications of Diversity.* New York: Verso, 2003. x, 180 p. Cloth $25.00.

MACINTYRE, GORDON: *Dugald Stewart: The Pride and Ornament of Scotland.* Portland, OR: Sussex, 2003. xii, 335 p. Paper $29.95.

MACK, MICHAEL: *German Idealism and the Jew: The Inner Anti-semitism of Philosophy and German Jewish Responses.* Chicago: University Press, 2003. viii, 229 p Cloth $35.00.

MACKLEM, TIMOTHY: *Beyond Comparison: Sex and Discrimination.* New York: Cambridge, 2003. x, 212 p. Cloth $65.00, paper $26.00.

MAGNANI, LORENZO: *Philosophy and Geometry: Theoretical and Historical Issues.* Boston. Kluwer, 2001. xix, 254 p. Paper $38.00.

MALACHOWSKI, ALAN: *Richard Rorty.* Princeton: University Press, 2002. xxi, 202 p. Cloth $55.00, paper $17.95.

MALCOLM, NOEL: *Aspects of Hobbes.* New York: Oxford, 2002. xii, 644 p. Cloth $49.95.

MALI, JOSEPH: *Mythistory: The Making of a Modern Historiography.* Chicago: University Press, 2003. xiii, 354 p. Cloth $40.00.

MANSBACH, ABRAHAM: *Beyond Subjectivism: Heidegger on Language and the Human Being.* Westport, CT: Greenwood, 2002. xiii, 158 p. Cloth $62.95.

MARENBON, JOHN: *Boethius.* New York: Oxford, 2003. xvi, 252 p. Cloth $55.00, paper $19.95.

MARGALIT, AVISHAI: *The Ethics of Memory.* Cambridge: Harvard, 2002. xi, 227 p. Cloth $24.95.

MARGOLIS, JOSEPH: *Reinventing Pragmatism: American Philosophy at the End of the Twentieth Century.* Ithaca, NY: Cornell, 2002. x, 179 p. Cloth $39.95; *The Unraveling of Scientism: American Philosophy at the End of the Twentieth Century.* Ithaca, NY: Cornell, 2003. xiii, 178 p. Cloth $35.00.

MARKELL, PATCHEN: *Bound by Recognition.* Princeton: University Press, 2003. xiii, 284 p. Cloth $59.50, paper $19.95.

MARTIN, JACK, JEFF SUGARMAN, and JANICE THOMPSON: *Psychology and the Question of Agency.* Albany: SUNY, 2003. x, 186 p. Cloth $57.50, paper $18.95.

MARTIN, MICHAEL: *Atheism, Morality, and Meaning.* New York: Prometheus, 2002. 330 p. Paper $21.00

MARTIN, THOMAS: *Oppression and the Human Condition: An Introduction to Sartrean Existentialism.* Lanham, MD: Rowman and Littlefield, 2002. xv, 155 p. Cloth $65.00, paper $19.95.

MATTHEWS, GERALD, MOSHE ZEIDNER, and RICHARD D. ROBERTS: *Emotional Intelligence: Science and Myth.* Cambridge: MIT, 2002. x, 697 p. Cloth $55.00.

MCCARTHY, GEORGE E.: *Classical Horizons: The Origins of Sociology in Ancient Greece.* Albany: SUNY, 2003. viii, 202 p. Cloth $65.50, paper $21.95.

MCEVILLEY, THOMAS: *The Shape of Ancient Thought.* New York: Allworth, 2002. xxxvii, 731 p. Cloth $35.00.

MCLAREN, MARGARET A.: *Feminism, Foucault, and Embodied Subjectivity.* Albany: SUNY, 2002. ix, 230 p. Cloth $65.50, paper $21.95.

MELE, ALFRED R.: *Motivation and Agency.* New York: Oxford, 2003. xiii, 264 p. Cloth $49.95.

METZINGER, THOMAS: *Being No One: The Self-model Theory of Subjectivity.* Cambridge: MIT, 2003. xii, 699 p. Cloth $55.00.

MILES, MURRAY: *Inroads: Paths in Ancient and Modern Western Philosophy.* Toronto: University Press, 2003. xxiv, 666 p. Cloth $95.00, paper $39.95.

MILKOV, NIKOLAY: *A Hundred Years of English Philosophy.* Boston: Kluwer, 2003. viii, 306 p. Cloth $125.00.

MILLER, ALEXANDER: *An Introduction to Contemporary Metaethics.* Malden, MA: Polity, 2003. xi, 316 p. Cloth $64.95, paper $29.95.

MILLER, WILLIAM: *Faking It.* New York: Cambridge, 2003. xi, 289 p. Cloth $30.00.

MINER, ROBERT C.: *Vico: Genealogist of Modernity.* Notre Dame: University Press, 2002. xvi, 215 p. Cloth $37.50.

MIRI, MRINAL: *Identity and the Moral Life.* New York: Oxford, 2003. xvii, 132 p. Cloth $25.00.

MITCHELL, SANDRA D.: *Biological Complexity and Integrative Pluralism* New York: Cambridge, 2003. xv, 244 p. Cloth $70.00, paper $26.00.

MODELL, ARNOLD H.: *Imagination and the Meaningful Brain.* Cambridge: MIT, 2003. xiv, 253 p. Cloth $32.95.

MOGGACH, DOUGLAS: *The Philosophy and Politics of Bruno Bauer.* New York. Cambridge, 2003. x, 290 p. Cloth $60.00.

MOHANTY, J.N.: *Between Two Worlds: East and West.* New York: Oxford, 2002. ix, 134 p. Cloth $19.95.

MOITRA, SHEFALI: *Feminist Thought: Androcentrism, Communication, and Objectivity.* New Dehli: Munshiram Manoharlal, 2002. x, 156 p. Cloth $22.50.

MOL, ANNEMARIE: *The Body Multiple: Ontology in Medical Practice.* Durham, NC: Duke, 2002. xii, 196 p. Cloth $59.95, paper $19.95.

MOORE, ANDREW: *Realism and Christian Faith: God, Grammar, and Meaning.* New York: Cambridge, 2003. xi, 269 p. Cloth $65.00, paper $24.00.

MORRIS, WARREN FREDERICK: *Escaping Alienation: A Philosophy of Alienation and Dealienation.* Lanham, MD: UP of America, 2002. xxix, 350 p. Paper $59.00.

MORRISS, PETER: *Power: A Philosophical Analysis* New York: Palgrave Macmillan, 2002. xlix, 277 p. Paper $24.95.

MOSS, LENNY: *What Genes Can't Do.* Cambridge: MIT, 2003. xx, 228 p. Cloth $34.95.

MULGAN, TIM: *The Demands of Consequentialism.* New York: Oxford, 2001. vi, 313 p. Cloth $49.95.

MURPHY, GREGORY L.: *The Big Book of Concepts.* Cambridge: MIT, 2002. 555 p. Cloth $45.00.

MURPHY, LIAM, and THOMAS NAGEL: *The Myth of Ownership: Taxes and Justice.* New York: Oxford, 2002. ix, 228 p. Cloth $25.00.

MURPHY, MARK C.: *An Essay on Divine Authority.* Ithaca, NY: Cornell, 2002. x, 198 p. Cloth $39.95.

MUTHU, SANKAR: *Enlightenment against Empire.* Princeton: University Press, 2003. xiv, 348 p. Cloth $69.50, paper $19.95.

NAAS, MICHAEL: *Taking on the Tradition: Jacques Derrida and the Legacies of Deconstruction.* Stanford: University Press, 2003. xxx, 211 p. Cloth $55.00, paper $21.95.

NADDAFF, RAMONA A.: *Exiling the Poets: The Production of Censorship in Plato's Republic.* Chicago: University Press, 2002. xv, 189 p. Cloth $27.50.

NADLER, STEVEN: *Spinoza's Heresy: Immortality and the Jewish Mind.* New York: Oxford, 2001. xi, 225 p. Cloth $35.00.

NAGEL, MECHTHILD: *Masking the Abject.* Lanham, MD: Lexington, 2002. ix, 128 p. Cloth $60.00, paper $24.95.

NAGEL, THOMAS: *Concealment and Exposure.* New York: Oxford, 2002. 240 p. Cloth $25.00.

NAILS, DEBRA: *The People of Plato. A Prosopography of Plato and Other Socratics.* Indianapolis: Hackett, 2002. xlviii, 414 p. Cloth $75.00.

NARVESON, JAN: *Respecting Persons in Theory and Practice: Essays on Moral and Political Philosophy.* Lanham, MD: Rowman and Littlefield, 2002. xiv, 313 p. Cloth $75.00, paper $24.95.

NEALE, STEPHEN: *Facing Facts.* New York: Oxford, 2001. xv, 254 p. Cloth $35.00.

NEIMAN, SUSAN: *Evil in Modern Thought: An Alternative History of Philosophy.* Princeton: University Press, 2002. xii, 358 p. Cloth $29.95.

NESHER, DAN: *On Truth and the Representation of Reality: A Collection of Inquiries from a Pragmatist Point of View.* Lanham, MD: UP of America, 2002. xxv, 465 p. Paper $58.00.

NEWMAN, ANDREW: *The Correspondence Theory of Truth: An Essay on the Metaphysics of Predication.* New York: Cambridge, 2002. xii, 251 p. Cloth $60.00.

NORRIS, CHRISTOPHER: *Hilary Putnam: Realism, Reason, and the Uses of Uncertainty.* New York: Manchester, 2002. viii, 280 p. Cloth $74.95, paper $29.95.

NORTON, BRYAN G.: *Searching for Sustainability: Interdisciplinary Essays in the Philosophy of Conservation Biology.* New York: Cambridge, 2003 viii, 554 p. Cloth $85.00, paper $30.00.

NUTTALL, JON: *An Introduction to Philosophy.* Malden, MA: Polity, 2002. xvi, 254 p. Cloth $64.95, paper $27.95.

O'CALLAGHAN, JOHN P.: *Thomist Realism and the Linguistic Turn: Toward a More Perfect Form of Existence.* Notre Dame: University Press, 2003. ix, 359 p. Cloth $59.95.

O'CONNER, PEG: *Oppression and Responsibility: A Wittgensteinian Approach to Social Practices and Moral Theory.* University Park: Penn State UP, 2002 xv, 152 p. Cloth $35.00.

O'GRADY, PAUL: *Relativism.* Montreal: McGill-Queen's, 2002. xi, 196 p. Cloth $70.00, paper $27.95.

OLLMAN, BERTELL: *Dance of the Dialectic: Steps in Marx's Method.* Urbana: Illinois UP, 2003. x, 234 p. Cloth $39.95, paper $18.95.

OLSCAMP, PAUL J.: *Moral Leadership: Ethics and the College Presidency.* Lanham, MD: Rowman and Littlefield, 2003. xiv, 169 p. Cloth $70.00, paper $26.95.

OLSON, CARL: *Indian Philosophers and Postmodern Thinkers: Dialogues on the Margins of Culture.* New York: Oxford, 2002. xv, 331 p. Cloth $35.00.

OLSON, PAUL A.: *The Kingdom of Science: Literary Utopianism and British Education, 1612–1870.* Lincoln: Nebraska UP, 2002. xii, 375 p. Cloth $65.00.

O'MANIQUE, JOHN: *The Origins of Justice: The Evolution of Morality, Human Rights, and Law.* Philadelphia: Pennsylvania UP, 2003. xii, 206 p. Cloth $36.50.

O'NEILL, ONORA: *Autonomy and Trust in Bioethics*. New York: Cambridge, 2002. xi, 213 p. Cloth $55.00, paper $20.00.

ORENSTEIN, ALEX: *W.V. Quine*. Princeton: University Press, 2002. ix, 209 p. Cloth $55.00, paper $17.95.

OTTESON, JAMES R.: *Adam Smith's Marketplace of Life*. New York: Cambridge, 2002. xiii, 338 p. Cloth $70.00, paper $26.00.

OVERALL, CHRISTINE: *Aging, Death, and Human Longevity*. Berkeley: California UP, 2003. xi, 264 p. Cloth $44.95, paper $29.95.

PANGLE, LORRAINE SMITH: *Aristotle and the Philosophy of Friendship*. New York: Cambridge, 2003. vii, 255 p. Cloth $65.00.

PAPINEAU, DAVID: *The Roots of Reason. Philosophical Essays on Rationality, Evolution, and Probability*. New York: Oxford, 2003. viii, 242 p. Cloth $29.95; *Thinking about Consciousness*. New York: Oxford, 2002. xiv, 266 p. Cloth $35.00.

PARKS, JENNIFER A.: *No Place Like Home? Feminist Ethics and Home Health Care*. Bloomington: Indiana UP, 2003. ix, 162 p. Cloth $29.95.

PENCE, GREGORY E.: *Brave New Bioethics*. Lanham, MD: Rowman and Littlefield, 2002. xiii, 177 p. Cloth $55.00, paper $17.95.

PEPERZAK, ADRIAN T.: *Elements of Ethics*. Stanford: University Press, 2004. xiv, 283 p. Cloth $60.00, paper $24.95.

PERRY, JOHN: *Identity, Personal Identity, and the Self*. Indianapolis: Hackett, 2002. xiv, 264 p. Cloth $39.95, paper $16.95.

PERRY, MICHAEL J.: *Under God? Religious Faith and Liberal Democracy*. New York: Cambridge, 2003. xv, 200 p. Cloth $60.00, paper $22.00.

PETTIT, PHILIP: *Rules, Reasons, and Norms*. New York: Oxford, 2003. xiii, 410 p. Paper $24.95.

PINCHES, CHARLES R.: *Theology and Action: After Theory in Christian Ethics*. Grand Rapids, MI: Eerdmans, 2002. x, 237 p. Paper $25.00.

PLANTINGA, ALVIN: *Essays in the Metaphysics of Modality*. MATTHEW DAVIDSON, ed. New York: Oxford, 2003. vi, 239 p. Cloth $60.00, paper $24.95.

PLUMWOOD, VAL: *Environmental Culture: The Ecological Crisis of Reason*. New York: Routledge, 2002. vii, 291 p. Cloth $75.00, paper $21.95.

POCOCK, J.G.A.: *Barbarism and Religion: The First Decline and Fall, Volume Three*. New York: Cambridge, 2003. xiii, 527 p. Cloth $60.00.

POPKIN, RICHARD H.: *The History of Scepticism: From Savonarola to Bayle*, Revised Edition. New York: Oxford, 2003. xxiv, 415 p. Cloth $74.00, paper $24.95.

POPKIN, RICHARD H., and AVRUM STROLL: *Skeptical Philosophy for Everyone*. Amherst, NY: Prometheus, 2002. 342 p. Cloth $35.00.

PRATT, SCOTT L.: *Native Pragmatism: Rethinking the Roots of American Philosophy*. Bloomington: Indiana UP, 2002. xviii, 316 p. Cloth $49.95, paper $21.95.

PRICKETT, STEPHEN: *Narrative, Religion, and Science: Fundamentalism versus Irony, 1700–1999*. New York: Cambridge, 2002. viii, 281 p. Cloth $65.00, paper $23.00.

PRIEST, GRAHAM: *Beyond the Limits of Thought*. New York: Oxford, 2002. xviii, 317 p. Cloth $65.00, paper $24.95.

PRINZ, JESSE J.: *Furnishing the Mind: Concepts and Their Perceptual Basis*. Cambridge: MIT, 2002. x, 358 p. Cloth $40.00.

PRIOR, ARTHUR N.: *Papers on Time and Tense.* PER HASLE, PETER ØHRSTRØM, TORBEN BRAÜNER, and JACK COPELAND, eds. New York: Oxford, 2003. ix, 331 p. Cloth $72.00, paper $24.95.

PUTNAM, HILARY: *The Collapse of the Fact/Value Dichotomy and Other Essays.* Cambridge: Harvard, 2002. ix, 190 p. Cloth $35.00.

RABINOW, PAUL: *Anthropos Today: Reflections on Modern Equipment.* Princeton: University Press, 2003. viii, 159 p. Cloth $45.00, paper $15.95.

RAEDER, LINDA C.: *John Stuart Mill and the Religion of Humanity.* Columbia: Missouri UP, 2002. xi, 402 p. Cloth $49.95.

RAILTON, PETER: *Facts, Values, and Norms.* New York: Cambridge, 2003. xx, 388 p. Cloth $80.00, paper $29.00.

REA, MICHAEL C.: *World without Design: The Ontological Consequences of Naturalism.* New York: Oxford, 2002. viii, 245 p. Cloth $35.00.

REISERT, JOSEPH R.: *Jean-Jacques Rousseau: A Friend of Virtue.* Ithaca, NY: Cornell, 2003. xiv, 211 p. Cloth $42.50.

REISS, TIMOTHY J.: *Mirages of the Self: Patterns of Personhood in Ancient and Early Modern Europe.* Stanford: University Press, 2003. xviii, 608 p. Cloth $70.00.

REMPEL, MORGAN: *Nietzsche, Psychohistory, and the Birth of Christianity.* Westport, CT: Greenwood, 2002. xi, 149 p. Cloth $62.95.

RESCHER, NICHOLAS: *Fairness: Theory and Practice of Distributive Justice.* New Brunswick, NJ: Transaction, 2002. xiii, 134 p. Cloth $34.95; *Epistemology: An Introduction to the Theory of Knowledge.* Albany: SUNY, 2003. xvii, 406 p. Cloth $81.50, paper $27.95, *On Leibniz.* Pittsburgh: University Press, 2003. x, 252 p. Cloth $42.50; *Sensible Decisions: Issues of Rational Decision in Personal Choice and Public Policy.* Lanham, MD: Rowman and Littlefield, 2003. ix, 147 p. Cloth $45.00.

RICCI, GABRIEL: *Time Consciousness: The Philosophical Uses of History.* New Brunswick, NJ: Transaction, 2002. xlvii, 192 p. Cloth $59.95.

ROACH, CATHERINE M.: *Mother/Nature: Popular Culture and Environmental Ethics.* Bloomington: Indiana UP, 2003. xvi, 221 p. Paper $19.95.

ROBERTS, ROBERT C.: *Emotions: An Essay in Aid of Moral Psychology.* New York: Cambridge, 2003. x, 357 p. Cloth $70.00, paper $25.00.

ROBINSON, DANIEL N.: *Praise and Blame: Moral Realism and Its Applications.* Princeton: University Press, 2002. xi, 225 p. Cloth $29.95.

ROBINSON, GUY: *Philosophy and Mystification: A Reflection on Nonsense and Clarity.* New York: Fordham, 2003. ix, 299 p. Paper $24.00.

ROCKMORE, TOM: *Marx after Marxism: The Philosophy of Karl Marx.* Malden, MA: Blackwell, 2002. xix, 224 p. Paper $24.95.

RODIN, DAVID: *War and Self-defense.* New York: Oxford, 2002. xvi, 213 p. Cloth $45.00.

RODRIGUEZ-PEREYRA, GONZALO: *Resemblance Nominalism: A Solution to the Problem of Universals.* New York: Oxford, 2002. xii, 238 p. Cloth $45.00.

ROSENBERG, JAY F.: *Thinking about Knowing.* New York: Oxford, 2002. viii, 257 p. Cloth $39.95.

ROUSE, JOSEPH: *How Scientific Practices Matter: Reclaiming Philosophical Naturalism.* Chicago: University Press, 2002. ix, 383 p. Cloth $49.00.

RUDD, ANTHONY: *Expressing the World: Skepticism, Wittgenstein, and Heidegger.* Chicago: Open Court, 2003. xii, 262 p. Paper $29.95.

RUSE, MICHAEL: *Darwin and Design: Does Evolution Have a Purpose?* Cambridge: Harvard, 2003. x, 370 p. Cloth $29.95.

SABAL, ANDREW: *Ruling Passions: Political Offices and Democratic Ethics.* Princeton: University Press, 2002. xv, 347 p. Cloth $69.50, paper $19.95.

SALLIS, JOHN: *On Translation.* Bloomington: Indiana UP, 2002. xii, 125 p. Paper $19.95.

SANTONI, RONALD E.: *Sartre on Violence: Curiously Ambivalent.* University Park: Penn State UP, 2003. xx, 179 p. Cloth $35.00.

SARKAR, HUSAIN: *Descartes' Cogito: Saved from the Great Shipwreck.* New York: Cambridge, 2003. xviii, 305 p. Cloth $65.00.

SCANLAN, JAMES P.: *Dostoevsky: The Thinker.* Ithaca, NY: Cornell, 2002. xiii, 251 p. Cloth $29.95.

SCANLON, T.M.: *The Difficulty of Tolerance: Essays in Political Philosophy.* New York: Cambridge, 2003. ix, 273 p. Cloth $55.00, paper $20.00.

SCHATZKI, THEODORE R.: *The Site of the Social: A Philosophical Account of the Constitution of Social Life and Change.* University Park: Penn State UP, 2002. xxii, 295 p. Cloth $45.00.

SCHICK, FREDERIC: *Ambiguity and Logic.* New York: Cambridge, 2003. ix, 154 p. Cloth $55.00, paper $19.00.

SCHMALTZ, TAD H.: *Cartesianism: The French Reception of Descartes.* New York: Cambridge, 2002. xiv, 288 p. Cloth $65.00.

SCHMITT, RICHARD: *Alienation and Freedom.* Boulder, CO: Westview, 2003. x, 145 p. Cloth $65.00, paper $19.00.

SCHOTT, ROBIN MAY: *Discovering Feminist Philosophy: Knowledge, Ethics, Politics.* Lanham, MD: Rowman and Littlefield, 2003. ix, 157 p. Cloth $65.00, paper $21.95.

SCHRADER-FRECHETTE, KRISTIN: *Environmental Justice: Creating Equality, Reclaiming Democracy.* New York: Oxford, 2002. xiii, 269 p. Cloth $35.00.

SCHRAG, CALVIN O.: *God as Otherwise Than Being: Toward a Semantics of the Gift.* Evanston, IL: Northwestern, 2002. xvi, 157 p. Cloth $79.95, paper $27.95.

SCHREIBER, SCOTT G.: *Aristotle on False Reasoning: Language and the World in the Sophistical Refutations.* Albany: SUNY, 2003. xv, 248 p. Cloth $68.50, paper $22.95.

SCHUELER, G. F.: *Reasons and Purposes: Human Rationality and the Teleological Explanation of Action.* New York: Oxford, 2003. xii, 174 p. Cloth $39.95.

SCOTT, CHARLES E.: *The Lives of Things.* Bloomington: Indiana UP, 2002. ix, 196 p. Cloth $49.95, paper $19.95.

SCOTT, PETER: *A Political Theology of Nature.* New York: Cambridge, 2003. xiii, 275 p. Cloth $65.00, paper $24.00.

SEARLE, JOHN R.: *Consciousness and Language.* New York: Cambridge, 2002. vii, 269 p. Cloth $65.00, paper $23.00.

SEDGWICK, EVE KOSOFSKY: *Touching Feeling: Affect, Pedagogy, Performativity.* Durham, NC: Duke, 2003. xii, 195 p. Paper $18.95.

SEN, AMARTYA: *Rationality and Freedom.* Cambridge: Harvard, 2002. ix, 736 p. Cloth $39.95.

SESSIONS, WILLIAM LAD: *Reading Hume's Dialogues: A Veneration for True Religion.* Bloomington: Indiana UP, 2002. x, 281 p. Cloth $49.95, paper $24.95.

SHAPIRO, GARY: *Archaeologies of Vision: Foucault and Nietzsche on Seeing and Saying.* Chicago: University Press, 2003. xxii, 437 p. Cloth $70.00, paper $27.50.

SHAPIRO, IAN: *The Moral Foundations of Politics.* New Haven, CT: Yale, 2003. xii, 289 p. Cloth $25.00; *The State of Democratic Theory.* Princeton: University Press, 2003. xi, 183 p. Cloth $19.95.

SHAPIRO, KAM: *Sovereign Nations, Carnal States.* Ithaca, NY: Cornell, 2003. ix, 186 p. Cloth $49.95, paper $19.95.

SHEEHAN, PAUL: *Modernism, Narrative, and Humanism.* New York: Cambridge, 2002. xiii, 234 p. Cloth $60.00.

SHERRATT, YVONNE: *Adorno's Positive Dialectic.* New York: Cambridge, 2002. xi, 254 p. Cloth $60.00.

SHIN, SUN-JOO: *The Iconic Logic of Peirce's Graphs.* Cambridge: MIT, 2002. x, 208 p. Cloth $34.95.

SHRAGE, LAURIE: *Abortion and Social Responsibility: Depolarizing the Debate.* New York: Oxford, 2003. xi, 173 p. Cloth $60.00, paper $17.95.

SHUSTERMAN, RICHARD: *Surface and Depth: Dialectics of Criticism and Culture.* Ithaca, NY: Cornell, 2002. xiv, 280 p. Cloth $49.95.

SIDERIS, LISA H.: *Environmental Ethics, Ecological Theology, and Natural Selection.* New York: Columbia, 2003. viii, 311 p. Cloth $52.50, paper $24.50.

SILVERMAN, ALLAN: *The Dialectic of Essence: A Study of Plato's Metaphysics.* Princeton: University Press, 2002. x, 393 p. Cloth $65.00, paper $24.95.

SINCLAIR, ALISON: *Uncovering the Mind: Unamuno, the Unknown and the Vicissitudes of the Self.* New York: Manchester, 2002. 248 p. Cloth $74.95, paper $24.95.

SINGER, MARCUS GEORGE: *The Ideal of a Rational Morality: Philosophical Compositions.* New York: Oxford, 2002. xix, 333 p. Cloth $60.00.

SISMONDO, SERGIO: *An Introduction to Science and Technology Studies.* Malden, MA: Blackwell, 2004. vii, 202 p. Paper $27.95.

SKINNER, QUENTIN: *Visions of Politics, Volume I: Regarding Method.* New York: Cambridge, 2002. xvi, 209 p. Cloth $65.00, paper $23.00; *Visions of Politics, Volume II: Renaissance Virtues.* New York: Cambridge, 2002. xix, 461 p. Cloth $65.00, paper $23.00; *Visions of Politics, Volume III: Hobbes and Civil Science.* New York: Cambridge, 2002. xvii, 386 p. Cloth $65.00, paper $23.00.

SLEINIS, E.E.: *Art and Freedom.* Urbana: Illinois UP, 2003. 237 p. Cloth $29.95.

SMITH, A.D.: *The Problem of Perception.* Cambridge: Harvard, 2002. ix, 324 p. Cloth $45.00.

SNODGRASS, JUDITH: *Presenting Japanese Buddhism to the West. Orientalism, Occidentalism, and the Columbian Exposition.* Chapel Hill: North Carolina UP, 2003. 351 p. Cloth $59.95, paper $21.50.

SOAMES, SCOTT: *Philosophical Analysis in the Twentieth Century, Volume I: The Dawn of Analysis.* Princeton: University Press, 2003. xix, 411 p. Cloth $35.00.

SOBLE, ALAN: *Pornography, Sex, and Feminism.* Amherst, NY: Prometheus, 2002. 228 p. Cloth $28.00.

SOLOMON, ROBERT C.: *Not Passion's Slave: Emotions and Choice.* New York: Oxford, 2003. xiv, 259 p. Cloth $35.00.

CAMBRIDGE

OUTSTANDING SCHOLARSHIP FROM CAMBRIDGE

Charles Taylor
Edited by Ruth Abbey
Contemporary Philosophy in Focus
$60 00 Hardback 0-521-80136-2 232pp
$20 00 Paperback 0-521-80522-8

Ancient Anger
Perspectives from Homer to Galen
Edited by Susanna Morton Braund and Glenn W. Most
Yale Classical Studies 32
$65 00 Hardback 0-521-82625-X 336pp

The Cambridge Companion to Abelard
Edited by Jeffrey E. Brower and Kevin Guilfoy
Cambridge Companions to Philosophy
$65 00 Hardback 0-521-77247-8 382pp
$25 00 Paperback 0-521-77596-5

John Stuart Mill
A Biography
Nicholas Capaldi
$40 00 Hardback 0-521-62024-4 464pp

Kant on Representation and Objectivity
A.B. Dickerson
$60 00 Hardback 0-521-83121-0 228pp

Self-Interest before Adam Smith
A Genealogy of Economic Science
Pierre Force
Ideas in Context 68
$65 00 Hardback 0-521-83060-5 296pp

The Cambridge Companion to Quine
Edited by Roger F. Gibson, Jr.
Cambridge Companions to Philosophy
$70 00 Hardback 0-521-63056-8 344pp
$25 00 Paperback 0-521-63949-2

Word and World
Practice and the Foundations of Language
Patricia Hanna and Bernard Harrison
$85 00 Hardback 0-521-82287-4 432pp
$30 00 Paperback 0-521-53744-4

*Prices subject to change

Kant: Prolegomena to Any Future Metaphysics
With Selections from the *Critique of Pure Reason*
Second Edition
Edited by Gary Hatfield
Cambridge Texts in the History of Philosophy
$50 00 Hardback 0-521-82824-4 269pp
$18 00 Paperback 0-521-53535-2

The Cambridge Companion to Brentano
Edited by Dale Jacquette
Cambridge Companions to Philosophy
$65 00 Hardback 0-521-80980-0 344pp
$24 00 Paperback 0-521-00765-8

Natural Kinds and Conceptual Change
Joseph LaPorte
Cambridge Studies in Philosophy and Biology
$70 00 Hardback 0-521-82599-7 237pp

Locke's Philosophy of Language
Walter R. Ott
$55 00 Hardback 0-521-83119-9 168pp

The Paradoxes of Art
A Phenomenological Investigation
Alan Paskow
$65 00 Hardback 0-521-82833-3 272pp

Hegel on Ethics and Politics
Edited by Robert B. Pippin and Otfried Höffe
The German Philosophical Tradition
$65 00 Hardback 0-521-81814-1 360pp

Can a Darwinian be a Christian?
The Relationship Between Science and Religion
Michael Ruse
$18 00* Paperback 0-521-63716-3 c 256pp
Now in paperback

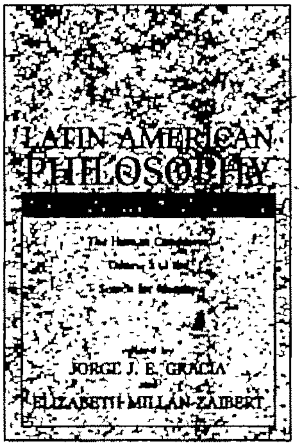
Descartes' *Meditations*
An Introduction
Catherine Wilson
Cambridge Introductions to Key Philosophical Texts
$55 00 Hardback 0-521-80981-9 284pp
$20 00 Paperback 0-521-00766-6

www.cambridge.org

CAMBRIDGE
UNIVERSITY PRESS

Prometheus Books

LATIN AMERICAN PHILOSOPHY FOR THE 21st CENTURY
The Human Condition, Values, and the Search for Identity
Edited by Jorge J. E. Gracia
and Elizabeth Millán-Zaibert
250 pp • ISBN 1-57392-978-6 • PB $28



EROS AND THE GOOD
Wisdom According to Nature
James S. Gouinlock
390 pp • ISBN 1-59102-148-0 • HC $32

THE IMPOSSIBILITY OF GOD
Edited by Michael Martin
and Ricki Monnier
438 pp • ISBN 1-59102-120-0 • HC $32

BEHOLD THE ANTICHRIST
Bentham on Religion
Delos B. McKown
366 pp • ISBN 1-59102-116-2 • HC $32

Humanity Books
an imprint of Prometheus Books

SKEPTICISM IN RENAISSANCE AND POST-RENAISSANCE THOUGHT
New Interpretations
Edited by José R. Maia Neto
and Richard H. Popkin
252 pp • ISBN 1-59102-156-1 • HC $65

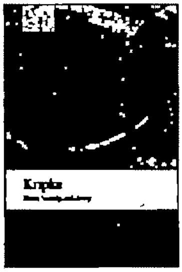
KROPOTKIN
The Politics of Community
Brian Morris
314 pp • ISBN 1-59102-158-8 • HC $50

SARTRE AND CAMUS
A Historic Confrontation
Edited and Translated
by David A. Sprintzen
and Adrian van den Hoven
275 pp • ISBN 1-59102-157-X • HC $45

For exam copies contact LRisio@prometheusbooks.com
Call toll free (800) 421-0351 • Fax (716) 691-0137 • www.prometheusbooks.com

New from

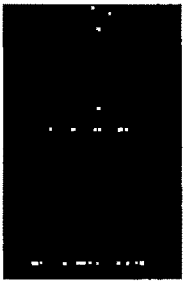
Oxford



Kripke
Names, Necessity, and Identity
CHRISTOPHER HUGHES

Saul Kripke, in a series of classic writings of the 1960s and 1970s, changed the face of metaphysics and philosophy of language. Christopher Hughes offers a careful exposition and critical analysis of Kripke's central ideas about names, necessity, and identity, and in the process makes significant contributions to continuing debates about such topics as modality, essence, natural kinds, and the relation between the mental and the physical.

April 2004 $49.95



New in Paperback!

Truth and the End of Inquiry
A Peircean Account of Truth
C. J. MISAK

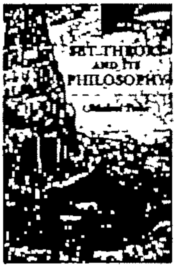
"The exposition and defense of Pierce's views on truth is clear, well organized, and largely nontechnical. It interacts with a good deal of the recent literature relevant to pragmatism and truth."
—*Choice*

paper April 2004 $26.95

Studies in the Philosophy of Logic and Knowledge
Edited by THOMAS BALDWIN AND TIMOTHY SMILEY

Questions about knowledge, and about the relation between logic and language, are at the heart of philosophy. Eleven distinguished philosophers from Britain and America contribute papers on such questions. The first half of the book constitutes a running debate about knowledge, evidence and doubt. The second half tackles questions about logic and its relation to language.

2004 $55.00



Set Theory and Its Philosophy
An Introduction
MICHAEL POTTER

Anyone wishing to work on the logical foundations of mathematics must understand set theory, which lies at its heart. What makes the book unique is that it interweaves a careful presentation of the technical material with a penetrating philosophical critique. Potter does not merely expound the theory dogmatically but at every stage discusses in detail the reasons that can be offered for believing it to be true.

April 2004 paper $29.95 cloth $74.00

The Divine Lawmaker
Lectures on Induction, Laws of Nature, and the Existence of God
JOHN FOSTER

John Foster begins by developing a solution to the problem of induction—a solution that involves the postulation of laws of nature, as forms of natural necessity. He then offers a radically new account of the nature of such laws and the distinctive kind of necessity they involve. Finally, he uses this account as the basis for an argument for the existence of God as the creator of the laws and the universe they govern.

2004 $39.95

Prices are subject to change and apply only in the US. To order or for more information, please call 1-800-451-7556. Visit our website at www.oup.com/us

OXFORD

REVUE INTERNATIONALE DE PHILOSOPHIE

Editor: Michel MEYER
Université de Bruxelles, C.P. 188
50, avenue F.D. Roosevelt - 1050 Brussels, Belgium

Each number is devoted to a particular movement, a particular philosopher, or a particular problem.

We publish 4 issues annually. Articles are written in English, French, German, or Italian.

Our last issue was devoted to *Dilthey.*

Articles by: Rudolf A. Makkreel: *Presentation* – Sylvie Mesure: *Individus et ensembles dans la méthodologie diltheyenne des sciences sociales* – Jos de Mul: *Das Schauspiel des Lebens: Wilhelm Dilthey and Historical Biography* – Frithjof Rodi: *Drei Bemerkungen zu Diltheys Aufsatz „Die Entstehung der Hermeneutik" von 1900* – Matthias Jung: *„Das Leben artikuliert sich" - Diltheys performativer Begriff der Bedeutung Artikulation als Fokus hermeneutischen Denkens* – Hans Ineichen: *„Die Entstehung der Hermeneutik" im Zusammenhang mit dem Spätwerk von Wilhelm Dilthey* – Jean Grondin: *La solution de Dilthey au problème du relativisme historique* – Tom Rockmore: *Dilthey and Historical Reason* – Rudolf A. Makkreel: *The Productive Force of History and Dilthey's Formation of the Historical World*

Coming numbers: Adorno – La pensée chinoise – L'Auto-organisation.

THE MONIST

AN INTERNATIONAL JOURNAL OF GENERAL PHILOSOPHICAL INQUIRY

JULY 2003 ◆ VOL. 86, NO. 3 ◆ SINGLE ISSUE, US $12

MORAL DISTANCE

ADVISORY EDITOR: DEEN K. CHATTERJEE

ARTICLES

Deen K. Chatterjee	*Moral Distance: Introduction*
Jeremy Waldron	*Who Is My Neighbor?: Humanity and Proximity*
Wendy C. Hamblet	*The Geography of Goodness: Proximity's Dilemma and the Difficulties of Moral Response to the Distant Sufferer*
Soran Reader	*Distance, Relationship and Moral Obligation*
Richard J. Arneson	*Consequentialism vs. Special-Ties Partiality*
Garrett Cullity	*Asking Too Much*
Jan Narveson	*We Don't Owe Them a Thing! A Tough-minded but Soft-hearted View of Aid to the Faraway Needy*
Kok-Chor Tan	*Patriotic Obligations*
Catherine Wilson	*A Humean Argument for Benevolence to Strangers*
Nicholas Rescher	*By the Standards of Their Day*
Stephen M. Gardiner	*The Pure Intergenerational Problem*
Karen Green	*Distance, Divided Responsibility and Universalizability*

Editor: Barry Smith, University at Buffalo, State University of New York
Single issues, $12 postpaid. Annual subscription: institutions, US $50; individuals, US $30; make checks payable to THE MONIST, 315 Fifth Street, Peru, IL U.S.A. 61354.
Website: http://monist.buffalo.edu E-mail: philomon1@netscape.net

---

THE MONIST

AN INTERNATIONAL JOURNAL OF GENERAL PHILOSOPHICAL INQUIRY

OCTOBER 2003 ◆ VOL. 86, NO. 4 ◆ SINGLE ISSUE, US $12

ART AND THE MIND

ADVISORY EDITORS: NICOLAS J. BULLOT & PASCAL LUDWIG

ARTICLES

Noël Carroll	*Art and Mood: Preliminary Notes and Conjectures*
Greg Currie & Jon Jureidini	*Art and Delusion*
Mark DeBellis	*Schenkerian Analysis and the Intelligent Listener*
Mark Rollins	*The Mind in Pictures: Perceptual Strategies and the Interpretation of Visual Art*
Dominic McIver Lopes	*Pictures and the Representational Mind*
Robert Hopkins	*Pictures, Phenomenology and Cognitive Science*
John Hyman	*Subjectivism in the Theory of Pictorial Art*

Editor: Barry Smith, University at Buffalo, State University of New York

Single issues, $12 postpaid. Annual subscription: institutions, US $50; individuals, US $30; make checks payable to THE MONIST, 315 Fifth Street, Peru, IL U.S.A. 61354.
Website: http://monist.buffalo.edu E-mail: philomon1@netscape.net

THE JOURNAL OF PHILOSOPHY

Celebrates 100 Years

Special Limited Time Offer

Full Volumes 1-99

1904-2002

$50.00 per Volume

All volumes are available for $50.00 plus $5.00 shipping and handling.

Send check or money order to:

The Journal of Philosophy
1150 Amsterdam Avenue, MC 4972
Columbia University
New York, NY 10027
(212) 666-4419
www.journalofphilosophy.org

Calcutta University Library
Campus

THE JOURNAL OF PHILOSOPHY

Calcutta University Library
ALIPORE CAMPUS

VOLUME CI, NUMBER 4
APRIL 2004

167 *The Irrationality of Unhappiness and the Paradox of Despair* Sarah Buss

COMMENTS AND CRITICISM

197 *On Floyd and Putnam on Wittgenstein on Gödel* Timothy Bays

BOOK REVIEWS

211 Candace Vogler: *Reasonably Vicious* Henry S. Richardson
216 NEW BOOKS
219 NEW BOOKS: PAPERBACK REPRINTS

Published by **The Journal of Philosophy, Inc.**

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Akeel Bilgrami, Arthur C. Danto, Kent Greenawalt, Patricia Kitcher, Philip Kitcher, Isaac Levi, Mary Mothersill, Philip Pettit, Carol Rovane, Achille C. Varzi. *Editor Emeritus*: Sidney Morgenbesser. *Consulting Editors*: David Albert, John Collins, James T. Higginbotham, Charles D. Parsons, Wilfried Sieg. *Managing Editor*: John Smylie.

THE JOURNAL OF PHILOSOPHY is owned and published by the Journal of Philosophy, Inc. *President*, Arthur C. Danto; *Vice President*, Akeel Bilgrami; *Secretary*, Daniel Shapiro; *Treasurer*, Barbara Gimbel; *Other Trustees*: Lee Bollinger, Leigh S. Cauman, Kent Greenawalt, Michael J. Mooney, Lynn Nesbit.

All communications to the Editors and Trustees and all manuscripts may be sent to John Smylie, Managing Editor, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, New York 10027. FAX: (212) 932-3721. You may also visit our website at: www.journalofphilosophy.org

THE JOURNAL OF PHILOSOPHY

2004

SUBSCRIPTIONS (12 issues)

Individuals	$35.00
Libraries and Institutions	$75.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$15.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are required.

Published monthly as of January 1977; typeset and printed by Capital City Press, Montpelier, VT.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL's terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

POSTMASTER: Periodical postage paid at New York, NY, and other mailing offices.
POSTMASTER: Send address changes to the *Journal of Philosophy* at MC 4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027.

© Copyright 2004 by the Journal of Philosophy, Inc. ISSN 0022-362X

THE JOURNAL OF PHILOSOPHY
VOLUME CI, NO. 4, APRIL 2004

ALIPORE CAMPUS

THE IRRATIONALITY OF UNHAPPINESS AND THE PARADOX OF DESPAIR*

Happiness is an end in itself. This is so obvious that even those who argue against the single-minded pursuit of happiness take it for granted that, all else being equal, we have good reason to refrain from behaving in ways that make us unhappy: we have good reason to avoid unhappiness because unhappiness is bad.

Even if, however, by its very nature, unhappiness just is the sort of thing that (all else being equal) any rational person would rather take in as small a dose as possible, one might still wonder whether something more can be said to illuminate this fact. In the pages that follow I pursue this possibility. Our natural desire to avoid unhappiness has the support of our reason, I argue, not only because unhappiness is an unpleasant psychological state, nor, more generally, because being unhappy is the opposite of "doing well." Our desire to avoid unhappiness is also justified by a *formal* principle of rationality; for unhappiness is itself a form of irrationality.

The relevant rational principle not only supports our preference for happiness over unhappiness; it also explains why we have good reason to care about our happiness in the way that we characteristically do. We value our own happiness not simply because it is happiness, but because it is *ours*.[1] Other people's happiness matters too, of course.

---

* I would like to thank Barbara Herman, Elijah Millgram, Martha Nussbaum, Connie Rosati, Angela Smith, and David Velleman for their comments on earlier drafts. I am especially indebted to Maggie Little for her patient and probing suggestions. The footnotes do not begin to indicate how much this article owes to her help.

[1] In this article I am thus not interested in what Stephen Darwall calls "self-concern": "self-concern...is only incidentally egocentric—its object being, unlike that of attitudes *de se*, the individual one is, rather than oneself as such...self-concern [is] an instance, in one's own case, of an attitude one can have in principle toward any individual, a concern that the moral point of view is thought to express equally toward everyone"—"Self-interest and Self-concern," *Social Philosophy and Policy*, XIV, 1 (Winter 1997): 158–78, here p. 160.

0022-362X/04/0104/167–96

But if we are honest, we have to admit that, to us, our own happiness (and that of our loved ones) matters more. The principle to which I call attention here sheds light on why we are rational to privilege ourselves in this way. It does not require us to regard our motives as universalizable, nor our selves as one among many. It implies that there are circumstances in which, independent of any other-regarding considerations whatsoever, we must, if we are to avoid irrationality, either engage in self-interested behavior aimed at reducing our unhappiness or revise our beliefs about who we are.

In making my case, I will describe a form of unhappiness—unhappiness as despair—and I will argue that unhappiness in this form has a paradoxical structure that explains why we have reason to try to avoid it. If, as I will also argue, despair is not only an important form of unhappiness, but a pervasive form as well, then the principle of rationality that is hostile to despair implies, more generally, that we ought to do what we can to avoid unhappiness.

At the heart of my account is an aspect of our self-concern that does not get as much attention as it deserves: our "personal ideals." The important role played in our lives by our personal ideals reflects the fact that they combine into one condition-to-be-realized our views about what is valuable and our views about who we are. I will devote much of this article to exploring the nature of our ideals and the implications of the way they integrate our values and aspirations with our self-conception. This exploration will uncover the deep connection between our ideals and our hopes. And this discovery will lead, in turn, to a discussion of what I call "the paradox of despair." To see why despair is paradoxical is to see why our reason opposes it. And with this insight, we will finally arrive at the supplementary justification of our self-regarding desire to avoid unhappiness.

Since we can cease being unhappy without becoming happy, reason's demand to avoid unhappiness is not, strictly speaking, a demand to pursue happiness as an end in itself. But it comes very close. After all, pursuing happiness is the natural way that we human beings flee unhappiness. So becoming happy is a means to becoming less unhappy in the same way, and to the same extent, that becoming sated is a means to becoming less hungry: just as being free from hunger and being sated would be the same thing if it were not for the fact that one could put an end to hunger by killing oneself, so too, being free from unhappiness and being happy would be the same thing if it were not for the fact that one could put an end to unhappiness by killing oneself—or killing all hope.

I. UNHAPPINESS AS THE FAILURE TO REALIZE PERSONAL IDEALS

Happiness and unhappiness mean many different things to different people. What interests me in this article is the familiar, widely shared conception of the experience of (dis)satisfaction, or (dis)contentment, with oneself and one's life. According to this conception, happiness and unhappiness are moods or emotions (not necessarily intense) directed at one's own condition. They are emotional states that pervade one's "being in the world."[2] When one is truly unhappy, a shadow—sometimes dark, sometimes faint—is cast over everything one encounters.

Happiness and unhappiness are not mere qualitative states. Each has a cognitive, as well as an affective, component. We are (un)happy *about* (or *with*) our condition; and (at least in most cases) we *believe* that our situation warrants the feeling of (dis)satisfaction it generates. A person can be (un)happy without perceiving the source of her (un)happiness. But if she is happy, she must believe that "things are good." And if she is unhappy, she must believe that, somehow, "something is wrong."

We are naturally disposed to care about whether we experience our own lives in this way.[3] Though we may rarely think about whether we are happy, and though happiness is rarely the express object of our actions, a human being who is indifferent to her own happiness is a specimen as strange as she is rare. This fundamental mode of self-interest is distinct from our more general concern to live good lives. Our poor health is especially intolerable to us when it "gets us down"; and unhappiness can be provoked by events that do not really prevent us from living perfectly healthy, fulfilling lives.

The unhappy person's judgment that there is "something wrong" with her life is often difficult, if not impossible, for her to flesh out with a more determinate description. Nonetheless, whatever the details, unhappiness is distinct from many other forms of dissatisfaction. When someone is unhappy, it is not simply that some of her desires are unsatisfied—or even that she would prefer it if things were different in certain respects. (Surely, no human being with a working mind could fail to satisfy *these* conditions!) Nor is unhappiness essentially a matter of wanting something *very intensely*—the fancy car in the local lot, a roof over one's head. Rather, the unhappy person's judgment that "something is wrong" is the judgment that her life falls short of what counts as a good life for her—that she is not living the life her personal ideals mark out as "to-be-realized" by her. A person is happy, in the relevant sense, if and only if she has a positive feeling about the gap between her reality as she perceives it and her personal ideals. She is unhappy if and only if she has a negative feeling about the perceived gap. It may, in principle, be possible for someone to believe that no gap exists, and yet feel nothing in response. So, too, someone may be unmoved by even a very large perceived gap. But while these are important conceptual possibilities, they are rare human conditions. In almost every human being, the cognitive component of unhappiness naturally gives rise to the affective component.

---

[2] Clearly, then, the psychological state that interests me in this paper is not the purely formal conception of a final end which Aristotle seems to have in mind when he announces that all human beings want to be happy.

[3] Even if it is better to be Socrates dissatisfied than a pig satisfied, to be Socrates satisfied is best of all—or so we (and Socrates) think.

Because there is no principled limit to the content of our personal ideals, human beings can be unhappy about anything and everything—just as they can be frightened by anything and everything, and just as they can find anything and everything funny, or depressing, or boring. But whatever the particular object of one's unhappiness may be, it makes one unhappy insofar as, and only insofar as, one takes it to fall short of one's current personal ideals. To be unhappy is to feel that things are not, in some respect anyway—perhaps in a way one cannot yet identify—as they must be if one's life is to be a truly good life.

Some human beings are more susceptible to unhappiness than others because they are less tolerant of perceived discrepancies between the ideal and the real, or because they perceive discrepancies as being greater than they really are. The slightest discrepancy evokes misery in some; others are not so easily bothered—whether because they do not notice the discrepancy (or its extent), because they are more optimistic about the prospects of improvement, or simply because they are willing to settle for less. In any case, our happiness is linked to our personal ideals. So, if we are to understand our interest in our own happiness—and if, in particular, we are to understand what distinguishes unhappiness from the standard human condition of living with unsatisfied preferences—then we had better take a closer look at the role these ideals play in our lives.

Most of us take an approving attitude toward many things. There are many things, people, and deeds that we like, admire, judge good. Of these, however, we identify very few as part of our own good. This is because very few strike us as having anything very important to do with us; and this, in turn, is for two compatible and overlapping reasons: (1) they do not matter that much to us, and (2) they are not related closely enough to who we think we are.

We may be overwhelmed with admiration for a person's knowledge of botany, or for a master painting in the Louvre, or for a daring trek to the top of Mt. Everest, without having the least interest in acquiring or doing these good things ourselves. We may value the talents of Michael Jordan without having the least desire to "be like Mike." The point is not that we do not believe that our lives would be better if they contained these good things. To be sure, we might refuse the offer to climb Mt. Everest, even with all expenses paid. But if a Genie asked us whether we would like to know more about plants, or whether we would like to be able to dunk a basketball, we might not hesitate to accept her offer. The point is that, however wonderful we think it would be to have these wonderful things, we do not regard them as to-be-had *by us*. Like the ability to fly, they are very nice, indeed. But they are not essential—or even important—components of our own good. They are not essential to our flourishing. Our lives do not "fall short" because they lack these good things.[4]

The importance of distinguishing between the judgment that something is good and the judgment that something is part of *one's own* good is implicitly acknowledged in the many philosophical attempts to sketch "the good life" for human beings: not every good thing is essential, or even important, to a good human life. Having acknowledged as much, however, we need to go further. For there is more to a good human life than its relation to the human nature we all share; though some human goods are surely universal, not all of them are. Human beings come in many varieties. They differ greatly from one another with respect to their values and their self-conceptions. And so, they differ greatly with respect to the sort of lives that are good lives-for-them, and the sort of lives they take to be good lives-for-them. In short, human beings have different personal ideals; and so, they differ with respect to the conditions under which it is possible for them to be happy. The point is not that no two human beings could possibly have the same personal ideals: our personal ideals are "personal" not in the sense that they are unique to us, but in the sense that they represent *our* conception of the good life *for us*.

As we grow into self-consciousness, who/what we take ourselves to be and what changes we think it is reasonable to expect in our lives are mutually determining. And each, in turn, contributes to, and is constituted by, our personal ideals. From a very young age, we each begin constructing our own (rather fuzzy) picture of who we are and what we are capable of doing and becoming. This involves figuring out not only what it is to be human, but what it is to be *me—this particular human* who differs from all others. The process takes place, moreover, even as we are learning what we like, what we do not like, and what is *worth* liking. Indeed, these two developments are mutually reinforcing: our evolving self-conception influences our conception of what is valuable, and our evolving sense of what is valuable influences our self-conception. The whole complex process generates a conception of our good in the form of a set of personal ideals, or conditions to-be-realized by us. Thus, for example, a young child who comes to think of herself as a "jock" also comes to value being able to run fast and throw far, and to expect that she will be able to run fast and throw far. At the same time, because these accomplishments are among her personal ideals, they contribute to how she thinks of herself; she regards herself as the sort of person who has these ideals; that is, she thinks of herself as a "jock."

---

[4] This is a frequently overlooked reason for rejecting the conception of happiness as the satisfaction of the greatest number of one's desires.

Many of the nice conditions that are not part of our good are conditions we could satisfy if we really wanted to; and many more are conditions we could at least come closer to satisfying. Other conditions, however, are not part of our good precisely because we do not think we can possibly satisfy them. More carefully, *we* do not *take* them to be part of our good because we do not believe that satisfying them is a live possibility for us; our satisfying them is not, we think, something we can reasonably expect to happen.[5] Goods that fall into this category are not conditions to-be-satisfied by us; and so, though we may fantasize about what it would be like to satisfy them, we cannot regard them as goals, or ends, to pursue. Surely, they are components of *someone's* good life; but this good life could not possibly be our own.[6]

As the examples show, many conditions which fail to qualify as part of our good for this second reason can also fail to qualify for the first reason. Indeed, we may come to regard them as not very important precisely because we have no expectation of satisfying them. (This may be the story behind many adjustments to handicaps.[7]) Some disqualified goods, however, may be conditions we think it would be well worth pursuing if there were any reason to expect that this would make a difference. The only reason we do not regard them as part of our good is because we do not believe they could possibly be part of our lives.

---

[5] If there is a *teleological* constraint on our good, then this is it: flying is not part of our good because we are not "made to fly."

[6] "Who," Pascal asks, "is unhappy at having only one mouth? And who is not unhappy at having only one eye? Probably no men ever ventured to mourn at not having three eyes. But anyone is inconsolable at having none"—*Pensees*, W.F. Trotter, trans., in Pascal's *Pensees and The Provincial Letters* (New York: Random House, 1941), #409, p. 130. The possibility at issue here refers to what there is any likelihood that we, or anyone else, could do. It does not refer to what is *conceptually* possible. For more on this point, see pp. 184–86 and note 24 of this article

[7] For one interesting relevant discussion, see Edward Dolnick, "Deafness as Culture," *The Atlantic Monthly* (September 1993): 37–51. According to many in the deaf community, Dolnick reports, "deafness is not a disability. Instead, many deaf people now proclaim they are a subculture like any other. They are simply a linguistic

As we will see, this does not mean that *alterations* in our conception of what is possible automatically lead to *alterations* in our conception of our good. (If our conception of our own good altered every time we discovered that we were unrealistically optimistic in our appraisal of our possibilities, then we would have no immediate knowledge of what unhappiness is.) What is more, our animal nature imposes limits on the extent to which our conception of our good can reflect our appraisal of our possibilities. (As I will have occasion to stress later,[8] if we found ourselves in circumstances that made it unrealistic to expect to satisfy our most basic needs, most of us would be incapable of concluding that these needs are really not so basic after all, that we do not have to satisfy them in order to thrive.) These qualifications notwithstanding, however, the connection between our perceived possibilities, our self-conception, and our conception of our good goes very deep. Whenever we think that our good is at odds with our possibilities, we experience a psychic pressure to find a new way of

---

minority (speaking American Sign Language) and are no more in need of a cure for their condition than are Haitians or Hispanics.... So strong is the feeling of cultural solidarity that many deaf parents cheer on discovering that their baby is deaf" (pp. 37–38).

For another interesting discussion of the human tendency to adjust one's sense of normalcy (one's self-conception), see Wilfrid Sheed, *In Love with Daylight: A Memoir of Recovery* (New York: Simon and Schuster, 1995), pp. 9–54. "Polio," Sheed says, "caused me to lose something quite irreplacable, something I would have sworn I couldn't live without" (p. 13). But despite having feared that "I'd go crazy if I got polio and had to give up baseball, and of course football—and *walking*" (p. 26), he discovered that "the speed of mental adjustment can be quite uncanny" (p. 19)—that "you get used to things incredibly quickly, and are ransacking the horizon for new pleasures almost before the old ones are out the door" (p. 46). "If," Sheed confesses, "you'd told me just six months before that the height of happiness would soon consist of leaning on a pair of crutches and inhaling deeply, I'd have woken up screaming" (p. 34). But "once you are used to your new condition, your imagination becomes free once again to rest or amuse itself; you stop scanning the skies for miracles, and life returns to an agreeably small scale of operations" (p. 49). Like other polio victims he met, Sheed never thought of himself as "handicapped."

Recently, both economists and psychologists have been very interested in the capacity human beings have to "adapt" to conditions they thought would be incompatible with their happiness. As economist George Loewenstein explains, "when any event occurs to us, we make it ordinary." In other words, we naturally adjust our conception of what counts as ordinary-for-us. See Jon Gertner, "The Futile Pursuit of Happiness," *The New York Times Magazine* (September 7, 2003): 44–47, 86, 90, especially p. 47.

[8] See pp. 195–96.

flourishing.[9] This, I will argue, is the pressure of our commitment to being noninstrumentally rational. It takes the form of our natural impulse to pursue our own happiness.

II. UNHAPPINESS AS DESPAIR

If most things are not part of a person's good, and if, nonetheless, it is possible to be unhappy about anything, then there is one obvious sense in which unhappiness can be irrational: someone's unhappiness is irrational if her personal ideals are irrational. This sort of irrationality occurs when someone has a false conception of what counts as "her life," or when she is confused about what counts as a candidate human good. The borderlines are fuzzy, of course, but someone can, I think, plausibly be deemed irrational if she is made unhappy by the recognition that somewhere, someone has failed an exam, or if she is unhappy that the sky is blue, or that she is in excellent health—where these states of affairs have no connection to anything else that could justifiably make her unhappy.

To the extent that someone's unhappiness thus depends on perverse ideals, she has an obvious reason to do something that is likely to make her happier, namely, change her ideals. But this reason applies only to perverse cases of unhappiness. And more importantly, it is not a reason to be concerned about the fact that one is unhappy, or to be moved by an interest in becoming less so. Accordingly, it does not shed light on the legitimacy of our natural self-concern.

The key to this supplementary justification is, I will argue, the structure that all cases of unhappiness have in common: the unhappy person is unhappy that the reality of her situation falls short of her personal ideals. When she believes, in addition, that this gap cannot be closed any further, then her unhappiness has the form of despair. I will soon argue that contrary to conventional wisdom, this form is the garden variety of unhappiness. For now, it suffices to point out that life affords countless examples—from the young ballplayer who realizes that she cannot make the team, to the graduate student who realizes that she cannot "make it" in her chosen profession; from the woman who finally admits to herself that she will never be able to have a child, to the woman who knows that nothing can bring her dead child back to life. The emotional states of such people can vary widely in intensity. Indeed, a person can be in despair without even being aware of this fact. I want to argue, however, that in all cases, unhappiness as despair is a form of irrationality. A human being in despair is not necessarily *criticizably* irrational, but she is irrational nonetheless.

---

[9] As I indicate in note 29, the inner conflict referred to here can be characterized in Stoic terms: one wants to conform one's aspirations to nature's design of the world, but to do so would be, it seems, to flout nature's design of oneself.

In order to understand this rather dark pronouncement, it is necessary to take a close look at the relation between despair and hope. In particular, it is necessary to consider the paradoxical nature of this relation. To this end, I want to call the reader's attention to Kant's discussion of regulative ideals. According to Kant, regulative ideals reflect the demands of reason. And unless we recognize this fact, we will find ourselves in the grip of a paradox: the hope which these ideals necessarily presuppose will itself give rise to despair.

Kant makes this point in discussing both science and morality. When we do science, we are trying to make sense of natural phenomena. But this means that we must assume that the world is intelligible to us. This, in turn, means that we must assume that nature is a systematic unity. And so, Kant concludes, we must believe that it is reasonable to hope that we will discover this unity, and the unconditioned condition that underlies it.[10] Similarly, when we think about how we have reason to act, we necessarily presuppose that it is possible for us to achieve the only unconditional end: the Highest Good, happiness in proportion to virtue. If this end were not achievable, reason would not demand that we pursue it. So, insofar as we are committed to complying with reason's demands, we necessarily believe that it is reasonable for us to hope that the Highest Good will be realized.[11]

There seems to be a problem, however. For when we take stock of the evidence available to us, neither the scientific nor the moral hope

[10] "The law of reason which requires us to seek for this unity is a necessary law, since without it we should have no reason at all, and without reason no coherent employment of the understanding, and in the absence of this no sufficient criterion of empirical truth. In order, therefore, to secure an empirical criterion we have no option save to presuppose the systematic unity of nature as objectively valid and necessary"—Kant, *Critique of Pure Reason*, Norman Kemp Smith, trans. (New York: St. Martin's, 1965), A651/B679, p. 538.

[11] "It is in the view of reason, in the field of its theoretical employment, no less necessary to assume that everyone has ground to hope for happiness in the measure in which he has rendered himself by his conduct worthy of it. "—*Critique of Pure Reason*, A809/B837, p. 638.

In discussing the Highest Good, Paul Guyer calls attention to this basic feature which, according to Kant, every human ideal must have if it is to play a role in our choices. "Kant's position seems to be that while assurance that the noncontradictoriness of an end would be all that is needed from a theoretical point of view to make a course of action aimed at that end rational, human psychology is such that in fact it needs a greater incentive, a positive reason to believe its end is realizable"—"From a Practical Point of View: Kant's Conception of a Postulate of Pure Practical Reason," in *Kant on Freedom, Law, and Happiness* (New York: Cambridge, 2000), pp. 333–71, here p. 364.

seems to be justified: neither ideal seems to be realizable. No evidence could possibly support the claim that nature is rational; and surely it is far more probable that the natural world is *not* in perfect harmony with our own faculties. Things are at least as bad in the moral context. For there seems to be ample—indeed, overwhelming—evidence that happiness is not apportioned to virtue.[12] It thus seems that we have good reason to despair of ever realizing our ideals.

Yet if they really are our ideals, then we cannot but hope to realize them. So we seem to be stuck.[13] Again, it is only because we have certain ideals that we can despair of failing to realize them. So a necessary condition for the possibility of their being our ideals is a necessary condition for the possibility of our despair: our despair depends on our hope.

It might seem that one could escape the paradox by simply abandoning the ideals. But this, Kant argues, is not a genuine possibility. And once we see why this is so, we see how to resolve the paradox. It is not possible to abandon our regulative ideals because they are the product of reason itself; without them, there can be no scientific reasoning, and no rational action. This means that hope is a necessary condition for the possibility of reasoning about why things are as they are, and about what is worth doing. Since reason itself demands that we hope—since we cannot reason without hope—we are justified in hoping. In relation to this justification, the lack of supporting evidence is beside the point.

Kant's discussion of regulative ideals is, of course, far more nuanced than this crude sketch suggests. But for my purposes, it is enough to point out the light this discussion sheds on the nature of the ideals that especially interest me in this article—and on their relationship to hope and despair. Kant shows us that experiencing despair over the impossibility of realizing the regulative ideals of reason presupposes the hope that we will, indeed, realize them. Similarly, I believe, when we are unhappy at the thought that one of our personal ideals cannot be realized, this is precisely because we still hope to realize (or at least more nearly realize) it. Our despair at the unrealizability of our personal ideals presupposes that they "regulate" our choice of ends; and their playing this role presupposes the hope that we will, indeed, realize (or at least more nearly realize) them.[14]

---

[12] This evidence also appears to count against the assumption that the ends of nature form a systematic unity.

[13] Allen Wood puts the situation very nicely: "Just as the dialectic of pure theoretical reason produces an illusion which 'unceasingly mocks and torments' us in our pursuit of knowledge, so the pursuit of the ideal of pure practical reason will lead us necessarily into the troubled waters of illusion also, where we will be threatened with the unattainability of an ideal which we ourselves cannot establish, but with which we cannot cease to concern ourselves without forsaking the rationality which is proper to our own nature"—*Kant's Moral Religion* (Ithaca, NY: Cornell, 1970), p. 98.

The best way to underscore the Kantian lesson that despair requires hope is, I think, to consider cases. Take Jane, for example, who has spent many years studying classical ballet. She thinks of herself as a dancer, and in so doing, she adopts the personal ideal of being able to do all the things that a dancer must do. Having adopted this ideal—having come to regard dancing as to-be-realized by her—she reinforces her conception of herself as a dancer. And in adopting this self-conception, she thinks of herself as having the capacities of a dancer: if she works hard enough, she believes, she will be able to twirl and leap gracefully, and do all the other things a dancer must do.

She believes this. She really does. But lately, things have not been going so well. Jane refuses to deceive herself. "My dancing," she thinks, "is lousy. It falls short of what it must be if I am to be a dancer." This fact challenges the self-conception implicit in her personal ideal: it challenges her belief that she is a dancer, that dancing is a part of her good, that in dancing she realizes her potential—in short, that she can (with hard practice) do the things that a dancer must do.

Not only is Jane unhappy. She is at a crisis point in her life. It might seem that the obvious way for her to resolve the crisis is to give up the ideal of being a ballet dancer. But it is important not to underestimate the difficulty of doing so. Giving up a personal ideal involves renouncing something one values highly. Worse still, it involves abandoning the self with which one has identified in adopting this ideal. This can be extremely difficult—even impossible. After all, our self-conceptions and the identifications they imply are not wholly responsive to our will; we cannot simply *decide* to be someone else once we realize that we are mistaken about who we really are.[15] Not only, moreover, are such shifts in self-conception psychologically difficult; they are also in tension with the virtues of loyalty, faithfulness, and constancy, and of courageously "sticking up for what one believes in." A person thus has good (moral) reasons, as well as strong motives, for holding on to her ideals, and to the hopes they support.

---

[14] As Harry Frankfurt explains, a person "*guides* himself by reference to [his ideals]"—"On the Necessity of Ideals," in *Necessity, Volition, and Love* (New York: Cambridge, 1999), pp. 108–16, here p. 111. My point here is closely related to a familiar point about desires. As Alfred Mele explains, "If I am convinced that I cannot travel faster than the speed of light, or change the past, or defeat the current heavyweight champion of the world in a fair fight, then although I might wish that I could do these things, I do not *desire to do* them. Achieving the represented objects of action-desires is doxastically open for the agent: if she does not explicitly believe that she can A, at least she is not convinced that she cannot A. Any desire to A ('A' being an action variable), by its very nature, inclines the agent, in some measure, to A intentionally, or to try to A, or to try to put herself in a position to A. This is part of what it is to be an action-desire"—"Motivational Strength," *Noûs*, XXXII, 1 (1998): 22–36, here 25. My point is that not only are our "action-desires" sensitive to our beliefs about what is possible, but that insofar as our ideals are not mere wishes, but actually *guide* our action-desires, they are under pressure to conform to our expectations. It is not (as I will soon have occasion to stress) that they cannot possibly defy our expectations, but that insofar as they do so, we have an incoherent/conflicted self-conception.

Understandable, and even justifiable, though Jane's hope may be, the evidence clearly indicates that it is groundless. Because Jane knows this, she despairs. She despairs of ever being the dancer she still hopes she might be. She despairs because she still believes that being a dancer is part of her good, because she still has the self-conception which is linked to this conception of her good, because she still accepts the possibilities implicit in this self-conception, and because she clings to these possibilities as a basis for hope. Were she to cease hoping—were she, that is, to resign herself to her limitations as a dancer—she would not cease to value dancing. Nor would she turn down a Genie's offer to give her the capacity she now lacks. She would, however, no longer regard herself as "falling short." She would be like those of us who would love to have wings, but who understand that to fly is simply not a part of our good. Or like the blind person who never thinks of herself as a sighted person who cannot see. She would no longer be in despair.

The reader who reflects on the histories of human lives—real and imagined—should have little difficulty discovering the paradox of despair on both a large and small scale. I offer a second example of my own in the form of a contrast between two aging women whose youthful beauty has been an important part of their self-conception. Martha looks in the mirror and sees that her face has begun to sag. "I am not as beautiful as I once was," she thinks, with a regretful, but resigned, sigh. Mary looks in the mirror and she, too, sees that her face has begun to sag. "I am not as beautiful as I once was," she thinks with alarm. Neither Martha nor Mary has lost her ability to appreciate human beauty; neither has altered her aesthetic judgments. But whereas Martha believes that youthful beauty is no longer a possibility for her, and so ceases to regard it as to-be-realized by her,[16] Mary believes she can surely look much better than *this*, and makes an appointment with the hairdresser, the make-up artist, the plastic surgeon. She has seen what these wizards can do; so, though she is not happy, she has reason to hope.

---

[15] Note that the impossibility referred to here does not rule out the conceptual possibility discussed in note 24.

[16] Alternatively, we could imagine her as an instance of a common phenomenon to which Sheed calls our attention: "The physical decrepitude of old age seems to bother people surprisingly little. When they were young, the old-timers around here must have dreaded the thought of someday looking the way they do now, all scales and wattles, but when it comes, it must seem like the only way to look"—*In Love with Daylight*, p. 239.

Time goes by. Mary has now tried everything she can think of, and she still falls far short of her ideal of beauty. The evidence is overwhelming: she is not that type of human being any more; she does not "have it in her" to be (that) beautiful; being (that) beautiful is not a possible way for her to flourish; it is not part of her good; it cannot be for her a condition to-be-realized. Still, she is deeply committed to a self-conception which is incompatible with the evidence. In *identifying* herself *as* a beauty, she has *identified with* herself as a beauty. As far as she is concerned, there is no gap between who she is and who she takes herself to be. She thus refuses to give up hope. Unable to deceive herself about the evidence, she nonetheless continues to believe that it is still possible for her to be beautiful in the way only a young person can be. At the same time, she believes (she knows) that this is not possible. So she scrutinizes the mirror in a state of despair. Again, her despair is not simply a function of the fact that she *wants* to be beautiful, nor even of the fact that she wants this *very badly*, where this is a matter of placing it at the top of her preference rankings. She is in despair because she regards (her conception of) human beauty as a part of her good. It is, she believes, a component of her way of flourishing, and so, it is a way she might be—indeed, it is a way she can reasonably expect to be. The paradox is, of course, that she knows better—just as do those who refuse to give up hope despite having been repeatedly rebuffed by their beloved, despite having tried all the infertility treatments many times, despite having lost all enthusiasm for their job a long time ago.

When the conflict between a person's reality and her personal ideals is the result of her own limitations, her contradictory beliefs about her possibilities are essentially contradictory beliefs about who she is. It might seem that the conflict between the real and the ideal does not take this form when it is occasioned by the person's external circumstances. I believe, however, that no such neat contrast can be drawn; and this is because our circumstances are intricately entangled with our conception of who we are. To make this point vivid, I offer one last example. Consider a woman whose only child has just died. Though she believes that her little Johnny is gone forever, she simply cannot believe this. To resign herself to her child's death would be to resign herself to being someone who is not Johnny's mother—someone for whom being Johnny's mother is not a state-to-be-realized by her. So she cannot stop hoping that the whole thing is just a horrible nightmare. So she is condemned to despair.

Let me be clear: the misery in this case is occasioned by the child's death. More importantly, his mother is miserable *about* his death—not about her conflicting sense of who she is. There is, however, a structure to her misery. It is the structure of a conflict between two self-conceptions—a conflict between two conceptions of what is possible, two conceptions of what it is reasonable to expect. This is the structure of despair.

III. THE PERVASIVENESS OF DESPAIR

I believe that unhappiness as despair is a very common form of unhappiness. Indeed, I suspect that all unhappiness, from the most mild to the most intense, has the structure I have attributed to despair. To be unhappy is to be in despair—even if it is not always (or even usually) to be in despair's deepest depths.

This claim may well seem outrageous, vulnerable to powerful challenges from opposite sides. It may seem obvious that neither hope nor despair is a necessary condition of unhappiness. Surely, the alert reader will insist, there are many cases in which someone is in despair, and so is unhappy, precisely because she has given up hope; and surely, people are often unhappy, not because they believe they *cannot* realize some personal ideal, but simply because they believe they *have not* yet realized this ideal (as fully as they would like). I will try to respond to each challenge in turn.

First, then, my conception of despair appears to be vulnerable to any number of counterexamples in which a person is in despair precisely because she lacks all hope. Consider, for example, the case of someone who has just been forced at gun point to leave her burning home and walk for days until she has reached a refugee camp in another country, where she can barely get enough food to sustain herself. Can such a person not reasonably complain that her life falls far short of her ideals, even if she does not believe there to be any real possibility of realizing these ideals? According to this challenge, the possibility to which the refugee is committed in embracing the ideal of returning home and reestablishing a life in which she has plenty to eat is not a "live" possibility, but merely the conceptual possibility of living this way. So her despair at not being able to realize her ideal is not paradoxical, after all: she has no hope of ever "starting over" again.

In this account, the role I have assigned to the attitude of hope is replaced with something like a mere wish. I want to argue, however, that this substitution does not work: the negative condition constituted by an unfulfilled wish is not the negative experience of being unhappy; if someone merely *wishes* that a given state of affairs would obtain, then her negative attitude toward the status quo does not include the disappointed or frustrated *expectations* that are essential ingredients of unhappiness.

Let us take a closer look at the refugee's wish-like attitude toward her ideal. By stipulation, it reflects the belief that it is unreasonable to expect the ideal to be realized. Yet in order to inspire despair, this attitude cannot be an "idle" wish of the sort one might direct toward a state of affairs that one likes very, very much but believes cannot obtain outside the realm of fantasy. That is, it cannot be an idle wish, even though it must be a wish she believes can never be granted. But what sort of attitude could possibly satisfy this requirement? Only, I submit, a paradoxical one—only an attitude whose object is something one regards as to-be-achieved, even though one believes that it really can*not* be achieved. If one's belief that one cannot achieve X is a source of despair, this must be because X is one's end. And if X is one's end, then one must believe that it is possible to achieve it.

Despair would not be paradoxical if it were possible to be unhappy about one's failure to satisfy some desirable condition without regarding this condition as (more fully) realizable, and so, as an end. It might seem that nothing could be easier. After all, can one not believe that things ought to be a certain way, while recognizing that they cannot possibly be this way? The answer is that one can, indeed, do this, but that unless, in addition to believing that one's life ought to be a certain way, one also insists, or demands, that it be as it ought, one has insufficient psychological material for despair. One can be quite "philosophical"—even complacent, bored, or amused—at the recognition of yet another illustration of the fact that "ought to be" does not imply "can be." To be sure, one may be surprised that the two are out of sync in this particular case; one may wish that they were not. But such attitudes do not suffice for despair, as is evident from the fact that most people manage to "grow up" without being laid low by the inevitable discoveries of the many ways in which real life falls short of what it really ought to be.

Think, for a moment, of how many aspects of our world are not as you believe they should be. Surely, you do not like this fact. You would probably be willing to call it an "unhappy fact," and to express your view by saying "I am *not happy with* the way things are (in, for example, the slums of Calcutta)." But this attitude is not unhappiness. Your recognition of the gap between what is and what ought to be does not perturb the waters of your affective life; it is not the source of your "ups and downs." Or rather, when such gaps do "get you down," this is not simply because things are not as they ought to be, but because they are not as they *must* be, given your goals.[17]

Unhappiness comes on the scene only when one actually insists that reality (one's real life) be as it ought to be.[18] One may address one's demand exclusively to one's pillow at night, or—silently—to nothing, and no one, in particular. What matters is the demanding attitude itself. Unless, moreover, one believes that it is reasonable to expect one's demands to be satisfied, one cannot believe that it is reasonable to make them. So, since one necessarily takes it for granted that one's demands are reasonable (at least if one is both sincere and sane), it is part of the very point of one's demands that one is reasonable to expect them to be met. We thus arrive, again, at the paradox of despair: despair at the unrealizability of one's ideals presupposes the belief in their realizability; in despairing that one's demands cannot be met, even while believing that it is reasonable to make them, one reveals the hope at the heart of one's ideals, and so one reveals that one has not yet really given up hope after all.

It is important to stress, again, that the relevant demand that *X*, *Y*, or *Z* happen is not merely the claim that this is what "*justice* demands." Justice demands that something be done about the slums in Calcutta. But one can believe this—believe it fervently—without feeling unhappy. In order to be unhappy about something one does not like, one must not only prefer that it be otherwise, but this change must be something one aims at; it must not only appear to be desirable; one must embrace it as a goal to-be-achieved—not just someone else's goal, but one's own. Unhappiness is essentially paradoxical precisely because it is essentially the experience of having a goal that one cannot see one's way to (coming closer to) achieving. Having the goal requires hope; seeing no way to achieve this goal—or even to come closer to achieving it—inspires despair. In most animals the force of instinct ensures that they have no goals they regard as impossible to achieve. We are made differently, however; and our capacity for unhappiness is one of the consequences.

---

[17] These are, of course, not the only possible attitudes along a continuum. In "feeling sorry" for another person, for example, one may experience an intense pain which is as distinct from feeling unhappy as is the pain of a sprained ankle. When the person for whom things are not going well is one's own child, one's sympathetic response can easily metamorphose into unhappiness. This is because one "identifies with" her. Her hopes are one's own, and—however foolish this may be—one has hopes (not just dreams) *for* her That is, among *one's own* possibilities are ways *she* can be: one is definitely *not* the mother of someone who flunks out of high school, or robs a bank!

[18] Again, this is Kant's point about regulative ideals As Susan Neiman explains, "The positing of an end is equivalent to a demand for its realization, and ideas of reason simply are ends"—*The Unity of Reason: Rereading Kant* (New York. Oxford, 1994), p. 69. "Regulative principles...[are] simultaneously ideas of and motives for the realization of, a certain possiblity" (p. 89). Kant calls them sources of "guidance"—*Critique of Pure Reason*, A827/B855, p. 649.

Many will grant that despair is closely related to hope. But according to conventional wisdom, the two attitudes cannot be simultaneous if they are directed toward the same thing: despair is either the loss of hope or the attitude that supplants hope when it is lost.[19] If I am right, the conventional wisdom is mistaken, for it fails to do justice to the difference between ceasing to believe that something desirable is possible and experiencing this shift as a loss. To experience the impossibility as a loss, one must still regard the unsatisfiable condition as a condition that is not only worth satisfying, but to-be-satisfied, an end, a goal, a live possibility even still. There is another bit of conventional wisdom that almost acknowledges this paradoxical attitude. "She simply cannot reconcile herself to her situation," we say. In other words, though she has the situation vividly, and painfully, before her mind's eye, some part of her refuses to regard it as an accurate picture of reality. (Consider the way that despair tickles the edges of panic in the thought of the frantic (*desperate*) parent who rushes about in search of a child lost at the zoo: "What cannot be might nonetheless be." It *cannot* be because it *must not* be. And the demand does not change when "might be" changes horribly to "is.")

Insofar, then, as our personal ideals are not merely the object of fantasy—insofar as they are not merely the focus of our most heartfelt wishes—they resemble our regulative ideals in presupposing hope. They presuppose the hope that they will be realized; for without this hope, they could not function as ideals—goods to be realized by us. This means that there is something intrinsically paradoxical about being unhappy about the unrealizability of one's ideals.

I believe that all unhappiness takes this form. For, again, I cannot see how there could be negative emotional states that deserve to be called "unhappiness" despite lacking the goal-oriented structure I have tried to describe; I cannot make sense of the idea that someone could be unhappy about something (rather than merely wishing it were otherwise) if her unhappiness had nothing to do with her expectations. Perhaps I am wrong about this. Perhaps the sorrow we feel when a loved one has died, for example, has an essentially passive component, as well as the active component I have been stressing here.[20] To defend this possibility, however, one cannot merely insist that true love *just is* an emotion or attitude which can be manifested by a grief that has nothing to do with dashed expectations and unrealized ideals. One must make sense of this unhappiness; one must explain how it is, essentially, a manifestation of love.

---

[19] In analyzing hope, J.P. Day succinctly expresses the conventional wisdom on the subject: "A cannot despair that P at the same time as he hopes that P..."—"Hope," *American Philosophical Quarterly*, VI (April 1969): 89–102, here p. 94.

Even if this explanatory burden can be met—even if there are aspects of a person's perceived good that are independent of her ideals, and so lack the practical structure of these ideals—our conception of our good is generally inseparable from our ideals. In most cases of unhappiness our unrealized good thus functions for us as a practical guide, something in relation to which we orient ourselves. Accordingly, paradox is not simply a feature of the occasional weak-willed act, and of the more frequent episodes of self-deception. Paradox is at the center of one of the most ordinary experiences of our emotional life.

But what about that second challenge to my account—the challenge from the other side? Even if unhappiness presupposes hope, why should we think that it presupposes despair? Surely, there are many cases in which a person is unhappy about gaps between the ideal and the real that, she thinks, are difficult, but not impossible, to close. Could anything be more obvious?

In fact, however, the appearances are misleading. Wherever we seem to see unhappiness without despair, we are really seeing either someone who wishes that things were different but would rather not do what it takes to change them, or someone who is in despair, after all. Let me explain.

Someone might wish that the gap between her personal ideals and her reality were not so big. Suppose, however, that she does nothing to close this gap. This could be for one of two reasons: either she believes that she *cannot* close it, or she does not *want* to close it, all things considered. Since in the former case, she despairs of closing it, it is the latter case that interests us here. If she does not want to close the gap, this must be because she believes that the cost of doing so is too great—that realizing (or more nearly realizing) the relevant ideals would involve foregoing other good things which, all things considered, it is better not to forego.[21] But if this is her situation, then she is not in fact unhappy, all things considered, about the fact that she has failed to make the effort necessary to close—or at least to narrow—the gap between her reality and her ideals.

---

[20] I am grateful to Maggie Little for pressing me on this point

[21] Of course, if it would, in fact, be relatively easy for her to bring reality into greater harmony with her ideals, then there is good reason to suspect that they are not really her ideals, after all. We are to imagine a case in which the costs are high enough to make it quite difficult to pay them, but not so high as to induce despair.

Of course, she may well feel unhappy about the fact that she finds herself in a situation in which it would be better not to make the effort. But if this is the case, then the object of her unhappiness must be the high price of closing the gap; for otherwise she would not be unhappy about the fact that the unrealized ideals are unrealized. If she is unhappy about the fact that her ideals are unrealized, then she must be unhappy about the fact that she would have had to pay a very high price to realize them. But this is the unhappiness of despair: she is unhappy that she cannot both realize her ideals and preserve the other good things whose sacrifice constitutes the high price she would have to pay in order to do so. In other words, she is unhappy about the fact that she cannot have two things at once, each of which is, she thinks, essential to her living a good life. She is unhappy because she does not believe that it is possible to realize this ideal condition.

IV. THE IRRATIONALITY OF UNHAPPINESS

Having discovered the paradoxical structure of unhappiness, we are in a position to recognize the principle of rationality that opposes it. When we are unhappy about the unrealizability of our ideals, we hold self-contradictory beliefs. And when we do this, we violate a fundamental principle of theoretical reason. Since our reason is hostile to contradictions, it tells us to stop despairing. It tells us to stop being unhappy—and to avoid future unhappiness, too.

The point is not that we are especially rational when we manifest our natural tendency to react positively to a perceived harmony between our reality and our ideals. To say that our reason supports our self-concern is *not* to say that it tells us to be preoccupied with ourselves, and with our own happiness, in particular. For all our reason cares, we can be so caught up in our "projects," so immersed in the business of realizing our ideals and maintaining the bare necessities of life, that we do not have a spare moment to consider the relationship between the way things are and the way we believe it would be ideal for them to be, nor even half a spare moment to ask ourselves how we feel about this relationship.[22] Only when we do, in fact, feel unhappy does our reason reinforce our natural desire to escape this unpleasant condition. And again, this does not mean that, *all things considered*, we ought to do what we can to escape it, but only that one of the things to consider is the fact that we cannot remain in this condition without contravening a formal principle of rationality. In short, we have *a* reason to try to be happy. Without the slightest allusion to other rational agents, or to the objective value of happiness, our reason supports our self-concern.

---

[22] I would like to thank Martha Nussbaum for forcing me to clarify this point.

There are several ways to obey reason's demand to cease holding contradictory beliefs about our personal ideals. One way is to give up the ideals, as human beings often do when they make the transition from the innocence of childhood to the experience of adulthood. Our ideals would not be worth much, however, if we could give them up whenever they proved "unrealistic." And as I noted earlier, *we* would not be worth much either if we lacked the loyalty, constancy, and courage that holding on to our ideals sometimes requires. This is why the hope at the heart of despair is not wholly irrational. Or perhaps we should say that there are good reasons for some irrational hopes, reasons that pure reason does not understand. (There is sometimes good reason to cry over spilled milk. Better: when what is spilled is very precious to us, there is often good reason to cry—and cry, and cry.)[23]

If someone were not made unhappy by the death of her child, we would think there was something terribly wrong with her. We would not normally express this thought by saying that she has sufficient reason to be unhappy. But we could: she has reason to be unhappy because he is dead; his being dead is a reason—a sufficient reason—for her to be as unhappy as a human being can possibly be. Implicit in this verdict is an ideal of humanity: human beings with "any feeling" would be devastated by this event; and "having feelings" is something we take to be part of any minimally decent human life.

For most of us, this ideal of humanity is a personal ideal. And most of us are incapable of failing to realize it: no matter how hard we might try, we could not purge ourselves of the wide range of emotions that characterize our daily experience. (See, however, the following discussion of resignation.) Given our ideals, we rightly regard the naturalness of our feelings as a fortunate fact. Nonetheless, our natural susceptibility to unhappiness makes us vulnerable to irrationality. Indeed, if I am right, it *is*, essentially, a natural susceptibility to irrationality.

---

[23] Derek Parfit declares: "Grief isn't irrational simply because it brings unhappiness. To the claim 'Your sorrow is fruitless,' Hume replied, 'very true, and for that very reason I am sorry' "—*Reasons and Persons* (New York: Oxford, 1984), p. 169. "We may," Parfit writes, "object to a world in which our loved ones are taken away, but if they are taken away, we do not want to fail to experience the fact, to register it as an evil." I hope it is clear that my agreement with Parfit is far greater than my disagreement. On my account, unhappiness is not irrational because it is fruitless. Rather, it is irrational because it includes the belief in the fruitlessness of hoping that things will be different. What is more, the irrationality at the heart of unhappiness is only *a* reason to avoid it. As I have stressed, there are often other, even better, reasons to be unhappy.

Human life is such that for every one of us (if we live long enough), there will be occasions on which we cannot realize the aforementioned ideal of humanity without having self-contradictory beliefs—without, that is, experiencing one of our other ideals as a good it is both reasonable and unreasonable to expect ourselves to realize. As I suggested earlier, on these and other occasions, self-contradiction is also the price we must pay for realizing the meta-ideal of being "true to our ideals." So, though on all such occasions we have good reason for reacting as we do, we always also have a reason *not* to react this way: we would thereby avoid self-contradiction. Since the ideal of rationality is another of our personal ideals, the pressure to "overcome" our humanity, and to "give up" ("give up on") our old conception of what counts as a good life for us, is pressure we impose on ourselves.

Fortunately, we need not always give up our ideals in order to satisfy reason's demands to avoid the self-contradiction of unhappiness. Sometimes we have another option: we can do something to prove to ourselves that the evidence against the possibility of realizing an ideal is not as conclusive as it seems—that we are not really unreasonable to expect the ideal to be realized. Under the pressure of theoretical reason, someone may finally screw up the courage to quit her high-powered job, and seek less prestigious (and less lucrative) employment that will leave her some time to devote to her poetry. The evidence against her ability to "make it" as a poet is, she thinks, overwhelming. And she has been dragging herself through her days with a barely acknowledged despair at ever being the poet she still believes she might be. It has taken a while. But since she cannot bear to abandon hope, she has finally come to see that her only rational option is to take the steps necessary to establish that the evidence against her has been misleading.

This woman is lucky: it is possible for her to do something to try to justify her ideal. Many others are not so fortunate, however. Short of deceiving herself, for example, there is nothing the aggrieved mother can do to convince herself that her child is not really dead. In her case, theoretical reason thus gives her no choice: it commands her to give up her hope. Again, it may be that she cannot, or will not, obey. She may rightly sense that despair is her most honorable option, that she would be a traitor to what matters most if she were any less unhappy and broken.[24] It is an interesting question under what conditions a despair that was initially justified becomes unreasonable; but I will not try to answer this question here. In most cases, it seems, the desperate hope gradually atrophies, to be replaced by the resignation that lies in the vast no-man's-land between happiness and unhappiness. There are other cases, however, in which the painful modification of one's ideals yields a joyful affirmation of one's new self-conception, and of the new life that goes with it. Think, for example, of someone who has despaired for years over his homosexuality, and who finally adjusts his ideals to fit the contours of his possibilities as he perceives them.[25] Sartre is right to note how often we cling to our self-conception as a convenient excuse. But he is wrong to suggest that someone is necessarily in bad faith if his sense of his possibilities reflects his belief that he is a type of human being whose good differs in important ways from the good of most others.

---

[24] An ideal generally combines several valued things into one. Accordingly, it is rare that we must abandon an ideal entirely in order to avoid irrationality. As Little has reminded me, we can usually revise our ideals in such a way that aspects of the old remain in the new. Thus, the would-be dancer might strive to become a choreographer; the wheelchair-bound man who had been training to race the mile might strive to become the world's fastest paraplegic.

Recently, some philosophers have suggested that we cannot betray our deepest values and ideals without destroying ourselves. Thus, according to Frankfurt, "Agamemnon at Aulis is destroyed by an inescapable conflict between two equally defining elements in his own nature: his love for his daughter and his love for the army he commands... When he is forced to sacrifice one of these, he is thereby forced to betray himself. Rarely, if ever, do tragedies of this sort have sequels. Since the volitional unity of the tragic hero has been irreparably ruptured, there is a sense in which the person he had been no longer exists"—"Autonomy, Necessity, and Love," in his *Necessity, Volition, and Love*, pp. 129–41, here p. 139, n. 8 Similarly, Christine Korsgaard argues that "to violate [the conceptions of yourself that are most important to you] is to lose your integrity and so your identity, and to no longer be who you are. That is, it is to no longer be able to think of yourself under the description under which you value yourself and find your life to be worth living and your actions to be worth undertaking. It is to be for all practical purposes dead or worse than dead"—*The Sources of Normativity* (New York: Cambridge, 1996), p. 102.

I think David Velleman is right to criticize such claims as obscuring the distinction between a person's self-conception and her identity (her "self")—between self-betrayal and suicide—see his "Identification and Identity," in Sarah Buss and Lee Overton, eds., *Contours of Agency* (Cambridge: MIT, 2002), pp. 97–100. A severe blow to one's self-conception may be worse than death, but it is not the same thing: one can suffer such blows without ceasing to exist. (Think, for example, of what happens to Lord Jim at the beginning of Conrad's novel—and of how he, the same Lord Jim, responds to having done what he knows that he—Lord Jim—would *never* do, and so could not possibly have done.)

An appreciation of the fact that we can survive profound changes in our self-conception should warn us against mistaking the *contingent*, though very real, limitations on our possibilities for *conceptual* limitations that we could not, even in principle, exceed without ceasing to be. In other words, though a person's good clearly depends on who and what she is, it is conceptually possible for a person to change so much that her good changes too. Indeed, for all we have reason to believe about the conditions of personal identity, a person could persist as a member of another species; and so, for all we know, she could, in principle, acquire the features that would make flying an essential component of her good. Aristotle is right when he reminds us that "no one [who is rational] chooses to possess the whole world if he has first to become someone else...; he wishes for this only on condition of being whatever he is"—*Nicomachean Ethics*, W.D. Ross, trans., in *The Basic Works of Aristotle*, Richard McKeon, ed. (New York: Random House, 1941), 1166A19, p. 1081. But it does not follow, as Nussbaum suggests in commenting on this passage, that what could possibly be a good for us is limited by our imagination Even if a life "was so remote from mine that I could not imagine in it a person whom I could accept as identical with myself"—"Aristotle on Human Nature and the Foundation of Ethics," in J.E.J. Altham and Ross Harrison, eds., *World, Mind, and Ethics: Essays on the Ethical Philosophy of Bernard Williams* (New York: Cambridge, 1995), pp. 86–131, here p. 91—it might nonetheless be a life I could live. If this possibility exceeds the limits of my imagination, then it is not relevant to my ideals, and so, as Korsgaard and Frankfurt remind us, it cannot constrain my choices; it is nothing I must, or can, be "true to"

V. RESIGNATION

The gay man who eventually embraces his homosexuality and the bereft mother who eventually "resigns" herself to her child's death illustrate two very different ways of adjusting to conditions that limit the sort of life one can live: whereas he no longer desires to be heterosexual, she would give anything to have her son back. Yet, I said, this is not a contrast between a happy and an unhappy person. Though the woman is anything but overjoyed about her new self-conception, she is resigned to it. That is, she has no hope that she will be someone's mother again, and so no despairing, disappointed hope—no despair, no disappointment, no unhappiness.

Resignation is the absence of hope without paradox. In some forms it is benign; in others it approaches pathology. In all it is a no-man's-land between happiness and unhappiness. Consider, again, the woman whose child has died. She may be resigned to the fact that she cannot drink water while singing an aria; she may be resigned to the fact that she experiences monthly pain and discomfort when she menstruates; and she may (eventually) be resigned to the fact that

---

in deciding what to do. We can concede all this, however, while at the same time acknowledging that I could, conceivably, acquire ideals in the future which would require such a great self-transformation that I cannot now imagine surviving it.

[25] The account offered here sheds light on a basic feature of happiness to which Charles Larmore calls our attention in a recent article: "We are never in a position to grasp in advance the full character of our good, even in its broad outline. As a result, our happiness includes not just the anticipated good we achieve, but also the unexpected good which happens to us"—"The Idea of a Life Plan," *Social Philosophy and Policy*, XVI, 1 (Winter 1999): 96–112, here p. 103

her child is dead. If her attitude is the same in all of these cases, it nonetheless takes different forms. To say that she is "resigned" to the first condition is to suggest that she accepts it—and the limits it imposes on her possibilities—without giving it any thought. This could also be her attitude toward the second condition; or—alternatively—she could frequently note how nice it would be if things were otherwise.[26] Finally, if she really does manage to resign herself to her child's death (and many in her situation never do), then this would almost certainly at least initially involve an emotional/psychological numbness in response to everything having anything to do with him, and with the awful event.

Without straining ordinary language we can, if we want, characterize this last form of resignation as a different sort of despair: despair without anguish—indeed, without affect of any kind. When it is directed toward only one dimension of one's life, it creates a barricade against the entrance of hope along this dimension. When it is directed toward one's life as a whole, it is a kind of death.

Suicide is an extreme response to the belief that there is "nothing to live for," one's hopes are hopeless. But one can kill one's *hopes* without ending the life of one's body. One can abandon hope altogether by ceasing to have any ideals. Someone in this condition is psychically/emotionally numb. She has no conception of her possibilities. So she has no conception of who she is. So she is dead to herself. Like all who are dead, she is invulnerable to disappointment. She is not happy, of course. But she is not unhappy either.[27]

---

[26] Notice how different this attitude is from the unhappiness of disappointed expectations. As this example reminds us, there would be far more misery in the world if Freud had been right—if, that is, little girls really did conceive of themselves as deformed boys. At some point in their lives half the human population would suffer in much the way that many homosexuals suffer in our society today. Perhaps there will come a time when we will be able to laugh at the thought that human beings must overcome painful disappointment in order to accept their erotic longings for others of the same sex, just as we can now laugh at the thought that some (most!) human beings must overcome painful disappointment in order to accept the fact that they do not have a penis.

[27] Someone who is profoundly depressed might seem to fit this description: she is too depressed to hope for anything because she is too depressed to care about anything. In most cases of serious clinical depression, however, the person is in such great pain that she desperately hopes the pain will go away. This motive for suicide is distinct from the "anti-motive" of having nothing to live for, though clearly, the latter is a source of the former, and the power of the former reinforces the latter. Following the next paragraph in the text, I call attention to how extremely difficult it is for us to reduce the intensity of certain basic forms of suffering by simply picking new targets for our hope.

Because hopeless despair is affectless despair, it is easy not to notice that one is in this condition. For a powerful description of this numb state of self-alienation, consider the following poem by Emily Dickinson: "There is a pain—so utter—/ It

Far better to be irrational than to lack the capacity for happiness. Far better to be unhappy. But when some measure of happiness can be purchased by killing off some part of oneself, the price may be worth it. There may be no other way to "get on with one's life." Of course, in most such cases, what is dead is not really gone. So there is usually room in such lives for a special hope. Someone whose child has died may eventually dare to conceive of herself as someone who can think again with pleasure of the many wonderful moments she spent with him. If, at the same time, she is convinced that this can never be, then her hope will be a new source of despair. If she holds on long enough, however, she may one day realize her ideal: she may succeed in sustaining the absence of any hope that things will be as they once were, without sustaining an absence of feeling.[28]

And what of the hungry, frightened refugee? Can she ever truly resign herself to her lot? It does seem that others in her situation have done just this. But I wonder whether such apparent resignation is not really just a case of biding one's time—waiting for things to get better. Waiting resignedly, perhaps. But waiting, just the same. Hoping.

The difficulty of giving up hope in such a case is not merely psychological; it is a function of the basic animal drive to thrive, and of the fact that there are limits to the conditions under which human animals can thrive—limits imposed by their need to eat and drink, by their sensitivity to pain, and so forth. Like all animals, we humans have the "built-in" goal to thrive in the ways that are characteristic of our species. And this goal is naturally part of our conception of who we are: I am someone who does not live on roots and leaves, who drinks water (or some other liquid), who sleeps daily. As my earlier allusion to blindness suggests, our self-conceptions can leave out many of the features that characterize a normal healthy human animal: one can conceive of oneself as the kind of human being who does not see, or whose legs do not move, or who does not make insulin. If, however, my happiness required me to embrace a conception of myself as the kind of human being who is perpetually cold and hungry, then it

---

swallows substance up—/ Then covers the Abyss with Trance—/ So Memory can step/ Around—across—upon it—/ As one within a Swoon—/ Goes safely—where an open eye/ Would drop Him—Bone by Bone," from *The Poems of Emily Dickinson*, Thomas H. Johnson, ed. (Cambridge: Harvard, 1979), Volume II, p 460.

[28] Emotions come and go. It is thus highly likely that someone who has suffered such a profound loss will become unhappy again for certain periods of time. Upon looking over a photo album, for example, she might relive her loss as though it were just yesterday; whatever she thought and felt when she first heard the news would "all come back again."

would probably be impossible for me to be happy. My animal nature would continue to remind me that I really do regard a full belly as part of my good. It would not permit me to give up the aim of eating a decent meal. It would keep me hoping; and so it would sentence me to despair.

Resignation is not alchemy: it cannot turn bad states of affairs into good ones. Nonetheless, it is often the necessary first step in revising one's conception of one's good.[29] And so, it is often the first step—a rationally justified step—on the way back to happiness. Again, most such revisions do not depend on concluding that the grapes were probably sour anyway; for, again, someone can readily admit that she would welcome a change which she no longer hopes for—and even a change which she no longer takes to be very important. In some cases, however, when, like the transformed homosexual, a person concludes that he was wrong about his own good, the alteration in his personal ideals is inseparable from an alteration in his more general views about what is desirable. Suppose, to take another example, that Bob has believed since childhood that no human life is a good human life unless the human being who lives it has been "saved." And suppose that he despairs of ever achieving salvation himself. One way for him to alleviate his unhappiness would be for him to abandon his belief. He might not be able to do this, of course. But what if he could? What might this abandonment look like? He might conclude that there is no God, after all. Or he might conclude that he, at any rate, has no God. In either case, if he succeeds in becoming a nonbeliever, he will no longer value salvation. Indeed, he probably would not even recognize it if it hit him on the head. (This is the difference between a true nonbeliever and someone who has turned his back on God in despair.[30])

---

[29] Spinoza, Rousseau, and many others have tried to impress upon us how much our happiness depends on our ability to recognize and accept what we cannot change. This article can be read, in large part, as an exploration of this Stoic insight, and an attempt to show that the connection between refusing to accept what is necessary and being unhappy is a more intimate connection than that between cause and effect. It is interesting to compare my remarks about resignation with the following comment by Allan Bloom: "Irony flourishes on the disproportion between the way things are and the way they should be while accepting the necessity of this disproportion. It is a classical style because the ancients did not expect that reality could become rational. .moderation, rather than being the expression of a timid or easygoing soul, was for them the expression of one who has overcome hope and therefore indignation.... In short, irony seems to presuppose the distinction between theory and practice..."—*Love and Friendship* (New York: Simon and Schuster, 1993), p. 193

[30] My claim about despair thus represents a challenge to the teaching that despair is the worst sin because it involves giving up on God.

One can sometimes deceive oneself into believing that one's conception of one's good life has really changed—and with it, one's conception of oneself as someone for whom it is good to live such a life. Theoretical reason gives mixed messages about such self-deception. On the one hand, it tells us not to permit our wishes to determine our conception of reality. On the other hand, it tells us to avoid self-contradiction. I suspect that if deceiving oneself about what one really deems valuable and important does not actually cause one to change one's mind, then this sort of self-deception is especially difficult to maintain. No sooner is the grape in one's mouth than one concedes that life was no good without it; one is overwhelmed with relief and joy when God finally shows His face.

It is often difficult, too, to deceive oneself about one's real possibilities. But I suspect that many happy people whose ideals are far from being realized owe their happiness, at least in part, to their ability to induce in themselves the false belief that the evidence does really warrant their hope. In some of these cases, the self-deceived belief becomes self-fulfilling. In all cases, it is an antidote to despair.[31]

When we comply with reason's demand by deceiving ourselves in this second way—and especially when we are incapable of pursuing the other ways of complying—we are, it seems to me, in a condition that is as close to Kantian "rational faith" as it is possible to come in relation to one's personal ideals: we have embraced a "useful fiction" which has the support of a fundamental principle of reason, even though the evidence tells strongly against it. In this case, of course, reason's recommendation is conditional, in part, on contingent psychological factors; reason is not, itself, the source of the ideals. But nor is it, on the Kantian account, the source of the human desire for happiness, without which happiness would not be part of the Highest

---

[31] Sheed offers an autobiographical example of how deceiving oneself can contribute to the gradual process of reconceptualizing oneself in the way one must in order to be happy: "Nobody ever got used to blindness, or living in a wheelchair, or crawling to the bathroom, in his mind, yet when the time comes, the body often moves amazingly fast, handing out instructions as it goes on how to handle the latest disaster, like the next clue in a game. Your task is to keep your mind occupied between clues, keeping the truth or at least the words at bay until your body has learned the next moves. If, for instance, I'd been told at the beginning exactly how far I'd get with my exercises, and where I'd end up, I might indeed have felt like a cripple and quite bowed down by it, but I was able to convince myself that I was going all the way until my body was ready to tell me otherwise, and until I'd learned slowly for myself, pushing against that unyielding hand, that this is where you get off.... [A] firm belief in miracles and the power of prayer, and a trusting nature, did treat me to three and a half years of unbridled optimism, at the end of which I was so used to polio that I could barely remember what life had felt like without it"—*In Love with Daylight*, pp. 31–32.

Good. If, as I have suggested, our happiness is tied to our perception of the relationship between our actual condition and our personal ideals, then the useful fiction that the gap between our real life and our ideal life is not so big—or if big, then at least not so difficult to close—may play an important role in sustaining our rationally mandated faith in the possibility that someday everyone who ought to be happy will be.

VI. THE RATIONALITY OF OPPRESSION

Many arguments with other rational agents about what each has reason to expect of the others are, in a broad sense, *political* arguments. Before bringing this article to a close, I want to say a few words about the political implications of the irrationality of unhappiness. Let me begin by stressing what my thesis does *not* imply. It does *not* imply that you can deprive me of something essential to my good—that, for example, you can deny me the right to vote—and then declare that I am irrational to protest, since what you have taken away is no longer attainable. I can reasonably retort that it is *not* irrational for me to protest, since it (the right) *is* attainable; it is attainable, since you can easily change things so as to enable me to vote. In short, you can deny me the right to vote without forcing me to conclude that it is unreasonable for me to expect that you will grant me this right.

This having been said, your behavior (and that of your associates in power) may well convince me that I have no reason to expect that you will grant me the right any time *soon*. In this case, if I continue to demand the right *now*, and if this is not simply a matter of strategy (or rhetoric), I will almost surely be unhappy. ("*Almost* surely": unhappiness *requires* a tension in one's beliefs about what is possible; but again, though it is our natural way of experiencing this tension, I do not see any reason to insist that it is the inevitable consequence.) If I am unhappy (as opposed to merely judging that things are not as I would like them to be, nor as they should be), I will be (in the sense that we have been exploring) irrational: since I do not believe that it is possible to gain the right soon, I would be more rational if, rather than clutching at a desperate hope, I did not "let it bother me" that I will have to wait.

Again, this merely implies that I have *a reason* to cheer up—not that, *all things considered*, I *ought* to do so. Happiness is not the only thing of importance in life. Certainly, rationality is not. Furthermore, in changing my attitude, I need not alter my long-term expectations. And so I need not give up my belief that I will, eventually, be granted the right. Nor need I give up fighting for this right. Indeed, my hopeful long-term expectations will surely depend on my belief that I will never give up.

What if, however, your behavior (and that of those with whom you share power) convinces me that this belief, too, is unfounded—that there is no reason to hope I will ever be granted the right to vote (or to drive a car, or to attend university, or...)? Unfortunately, we are all too familiar with situations of this kind, and we know the hard choice that they offer: like so many other victims of oppression, I will have to choose between being unhappy for the rest of my life and being happy with less than I had hoped for. If I really have good reason to believe that these are my only alternatives where my happiness is concerned, and if it really is within my power to choose, then, it seems to me, I would not necessarily be wrong to opt for happiness. And, in any case, I would not be irrational to do so.

Of course, this is just what you, my oppressor, are counting on me to recognize. You are counting on me to try not to let my condition bother me (not to let it "get me down"), even if I dislike it; and you know that if it stops bothering me (if it stops "getting me down"), then I am eventually likely to take your restrictions for granted when I am thinking about who I am and what is essential to my own good—or if I will not ever reach this point myself, then my children will (or those of my sex, if the restrictions apply only to them). This outcome is compatible with my continuing to fight on behalf of the generations to come. But when I am no longer unhappy, the motivation to fight will be hard to sustain. If I see this subtler side to oppression clearly enough, I will recognize an additional reason to opt for unhappiness: I do not want to be a collaborator; I *am not* a collaborator.

Or am I? After all, I am certainly someone who wants to be happy. I hope that I will one day be happy. Can I really reconcile myself to regarding happiness as a very nice condition which is, however, not-to-be-achieved by me? And if I can, should I? What it is rational for me to demand of you (and of all other rational agents) depends on what it is rational for me to demand of myself. I may be confused about this. And since my confusion will make it more difficult for me to resist you, your success in oppressing me will in this case, too, be due, in part, to the irrationality of unhappiness.

There is considerable philosophical interest in the fact that we sometimes have *practical* reasons for holding beliefs that are not well supported by the evidence. In seeking a justification for our desire to be happy, I have discovered a simple *theoretical* reason for satisfying this desire.[32] It turns out that the impulse to bring our lives into line with our ideals and the desire to bolster our confidence in the appropriateness of these ideals are not really as distinct as we have been inclined to think.

With this discovery, we have, in effect, moved from Kant to Hegel. At least I think this is part of what Hegel was getting at when he told his magnificent story about how an ill-founded confidence in our conception of ourselves and our world generates a desire to *transform* ourselves and our world. Hegel to the contrary notwithstanding, the Real is not the Ideal. Nonetheless, we are committed to bringing them closer together. And we honor this commitment when we pursue our own happiness.

This is not to deny, of course, that we are also committed to acting from motives that we can justify to others. Our reason has heterogeneous ends; and nothing I have said by way of justifying our self-concern implies that this justification is ever overriding. It may well be that we should act on the desire to be happy only in those cases in which our behavior could be endorsed by all other rational agents. Nonetheless, if I am right, rational agents necessarily endorse the desire itself; this is part of what it is to be rational.

According to T.M. Scanlon, "individual well-being [is] morally significant...not because it is intrinsically valuable or because promoting it is self-evidently a right-making characteristic, but simply because an individual could reasonably reject a form of argument that gave his well-being no weight."[33] I have tried to show that one reason why arguments can reasonably be rejected if they give an individual's *happiness* no weight is precisely because, at least in most cases, unhappiness is intrinsically irrational; it is intrinsically irrational because it is itself a form of irrationality.

SARAH BUSS

The University of Iowa

---

[32] According to Darwall, "practical reason includes no intrinsic requirement that we care either about others or about ourselves, since rational agency seems possible without even the capacity to care about a person, oneself or another, for that person's sake"—"Self-interest and Self-concern," pp. 169–70. I am inclined to disagree. My point, however, is that even if Darwall is correct, reason in its *theoretical* capacity does support (at least one form of) our self-concern.

[33] "Contractualism and Utilitarianism," in Amartya Sen and Bernard Williams, eds., *Utilitarianism and Beyond* (New York: Cambridge, 1982), pp. 103–28, here p. 119

COMMENTS AND CRITICISM

ON FLOYD AND PUTNAM ON WITTGENSTEIN ON GÖDEL

In a recent discussion piece, Juliet Floyd and Hilary Putnam[1] present a new analysis of Wittgenstein's "notorious paragraph" on Gödel's first incompleteness theorem. Textually, they claim that Wittgenstein's remarks have been widely misunderstood, and they argue that Wittgenstein had a better understanding of Gödel's theorem than he has often been credited with. Substantively, they find in Wittgenstein's remarks "a philosophical claim of great interest," and they argue that, when this claim is properly assessed, it helps to vindicate some of Wittgenstein's broader views on Gödel's theorem.

Here, I address the second of these two arguments (while disclaiming the scholarly credentials required to assess the first, the purely textual, one).[2] I begin by examining the central claim which Floyd and Putnam attribute to Wittgenstein, showing that their argument for this claim is inadequate and that the claim itself is almost certainly false. I then argue that, even if Wittgenstein's central claim *were* true, it would not lead to the conclusions Floyd and Putnam think it does. At the end of the day, I conclude that Floyd and Putnam provide no new insights into Gödel's theorem by way of their reading of Wittgenstein.

I. ON FLOYD AND PUTNAM

Let us begin with the relevant passage from Wittgenstein.

> I imagine someone asking my advice; he says: "I have constructed a proposition (I will use 'P' to designate it) in Russell's symbolism, and by means of certain definitions and transformations it can be so interpreted that it says: 'P is not provable in Russell's system'. Must I not say that this proposition on the one hand is true, and on the other hand unprovable? For suppose it were false; then it is true that it is provable.

---

[1] "A Note on Wittgenstein's 'Notorious Paragraph' about the Gödel Theorem," this JOURNAL, XCVII, 11 (November 2000): 624–32

[2] For more on the textual questions, see Floyd, "On Saying What You Really Want to Say: Wittgenstein, Gödel, and the Trisection of the Angle," in Jaakko Hintikka, ed., *From Dedekind to Gödel: Essays on the Development of the Foundations of Mathematics* (Boston: Kluwer, 1995), pp. 373–425; Floyd, "Prose versus Proof: Wittgenstein on Gödel, Tarski, and Truth," *Philosophia Mathematica*, IX (2001): 280–307, and Mark Steiner, "Wittgenstein as His Own Worst Enemy: The Case of Gödel's Theorem," *Philosophia Mathematica*, IX (2001): 257–79

0022-362X/04/0104/197–210

> And that surely cannot be! And if it is proved, then it is proved that it is not provable. Thus it can only be true, but unprovable."
>
> Just as we can ask, "'Provable' in what system?," so we must also ask, "'True' in what system?" "True in Russell's system" means, as was said, proved in Russell's system, and "false" in Russell's system means the opposite has been proved in Russell's system.—Now, what does your "suppose it is false" mean? *In the Russell sense* it means, "suppose the opposite has been proved in Russell's system"; *if that is your assumption* you will now presumably give up the interpretation that it is unprovable. And by "this interpretation" I understand the translation into this English sentence.—If you assume that the proposition is provable in Russell's system, that means it is true *in the Russell sense*, and the interpretation "P is not provable" again has to be given up. If you assume that the proposition is true in the Russell sense, *the same* thing follows. Further: if the proposition is supposed to be false in some other than the Russell sense, then it does not contradict this for it to be proved in Russell's system. (What is called "losing" in chess may constitute winning in another game.)[3]

In this passage, we find Wittgenstein criticizing a relatively common interpretation of Gödel's first theorem: that the theorem shows—or helps to show—that there are true but unprovable sentences of ordinary number theory.[4] Wittgenstein objects to this interpretation, partially because he is skeptical concerning the notion of "Truth" in play here ("'True' in what system?"), and partially because he is opposed in principle to the derivation of "philosophical" claims from "mathematical" arguments.[5]

In their analysis of this passage, Floyd and Putnam focus on the following key claim (paraphrased from the second paragraph of Wittgenstein's discussion):

> (KC) If one assumes that $\neg P$ is provable, then one should give up the "translation" of P by the English sentence 'P is not provable'.

In what follows, I start by sketching the basic mathematics behind this claim. I then examine Floyd and Putnam's argument for the claim and what they take to follow from it. Finally, in sections II and III, I explain how and why their analysis goes wrong.

---

[3] *Remarks on the Foundations of Mathematics*, G.E. von Wright, R. Rees and G.E.M. Anscombe, eds., Anscombe, trans. (Cambridge: MIT, 1956), I, Appendix III, §8.

[4] Wittgenstein focuses on number theory as formulated in "Russell's system"—that is, the system of *Principia Mathematica*. There is, however, nothing in his argument that depends on this particular choice of background logic. For expository convenience, I will recast the argument in terms of ordinary, first-order Peano Arithmetic. Later—in section IV—I discuss the possible philosophical significance of formulating the argument in "Russell's system."

[5] See, for example, *Remarks*, VII §19. See also *Philosophical Investigations*, Anscombe, trans. (New York: Macmillan, 1953), §124.

I begin with the relevant mathematics. In his original paper on incompleteness,[6] Gödel defines two formulas in the language of formal number theory. These formulas—which I will call "PROOF(x, y)" and "SUBST(x, y, z)"—have the following nice property:[7]

> NUMERALWISE REPRESENTABILITY: Let m be the code of a sentence, ϕ, and let n be a natural number:
>
> (1) If n is the code of a proof of ϕ, then PA $\vdash$ PROOF$[\hat{n}, \hat{m}]$.
> (2) If n is not the code of a proof of ϕ, then PA $\vdash \neg$PROOF$[\hat{n}, \hat{m}]$.
>
> Similarly, let $\phi(v_0)$ be a formula with one free variable, let p be the code of ϕ, and let n and m be natural numbers:
>
> (1′) If m is the code of $\phi(\hat{n})$, then PA $\vdash$ SUBST$[\hat{p}, \hat{n}, \hat{m}]$.
> (2′) If m is not the code of $\phi(\hat{n})$, then PA $\vdash \neg$SUBST$[\hat{p}, \hat{n}, \hat{m}]$.

Using these formulas, Gödel defines the sentence which Wittgentein later calls "P" (although Gödel himself calls it something different). Initially, he defines the formula:

$$\psi(v_0) = \exists y\, [\mathrm{SUBST}(v_0, v_0, y)\ \&\ \neg\exists z\, \mathrm{PROOF}(z, y)]$$

Letting e_0 be the code for $\psi(v_0)$, Gödel sets:

$$P = \psi(\hat{e}_0) = \exists y\, [\mathrm{SUBST}(\hat{e}_0, \hat{e}_0, y)\ \&\ \neg\exists z\, \mathrm{PROOF}(z, y)]$$

This, then, gives the basic definition underlying Wittgenstein's (and later Floyd and Putnam's) discussion. There are two things we should notice about this definition.

First, the property of numeralwise representability explains why someone might think that P should be interpreted as saying "P is unprovable." On the assumption that PA is sound—that is, that $\mathbb{N} \models$ PA—the numeralwise representability of PROOF(x, y) and SUBST(x, y, z) entails the following, somewhat more semantic, property:[8]

---

[6] Gödel, "On Formally Undecidable Propositions of *Principia Mathematica* and Related Systems," in *Collected Works* (New York: Oxford, 1986), pp. 145–99.

[7] To unpack the terminology/notation here, I note that "coding" is simply a way of associating natural numbers to syntactic objects like formulas and proofs; done properly, it allows syntactic properties like "being well-formed" or "being a valid proof" to be treated as number-theoretic properties of the associated codes. Similarly, the notation $\hat{n}$ simply represents the expression for the number n in the language of formal number theory: specifically, $0 + 1 + \ldots + 1$, with n 1's.

[8] It is important to note that these facts about arithmetic expressibility can be proved directly, without passing through numeralwise representability (Gödel himself approaches arithmetic expressibility this way in his 1931 paper) Nonetheless, once we have numeralwise representability, it provides a quick route to arithmetic expressibility.

> ARITHMETIC EXPRESSIBILITY: For any two numbers n and m, $\mathbf{N} \models \mathrm{PROOF}(\bar{n}, \bar{m})$ if and only if m is the code of a sentence, ϕ, and n is the code of a proof of ϕ.
>
> Similarly, for any three numbers n, m, and r, $\mathbf{N} \models \mathrm{SUBST}(\bar{n}, \bar{m}, \bar{r})$ if and only if n is the code of a formula, $\phi(v_0)$, and r is the code of the sentence $\phi(\bar{m})$.

Using arithmetic expressibility, therefore, we can make perfect sense of the claims that $\neg\exists z\ \mathrm{PROOF}(z, y)$ "says that" the sentence coded by y is unprovable and that $\mathrm{SUBST}(\bar{k}_0, \bar{k}_0, y)$ "says that" y is the code of P. Together, these two claims provide the intuitive basis for interpreting P as "P is unprovable."

Second, our definition explains why someone might think that P is a true but unprovable sentence of number theory. If we filter P through arithmetic expressibility, we get the following result:

$$
\begin{aligned}
\mathbf{N} \models P\quad & \text{if and only if there is an } n\text{, such that } \mathbf{N} \models \mathrm{SUBST}(\bar{k}_0, \bar{k}_0, \bar{n}) \\
& \qquad \text{and } \mathbf{N} \models \neg\exists z\ \mathrm{PROOF}(z, \bar{n}). \\
& \text{if and only if there is an } n\text{, such that } n \text{ is the code of } P \\
& \qquad \text{and } \mathbf{N} \models \neg\exists z\ \mathrm{PROOF}(z, \bar{n}). \\
& \text{if and only if there is no } m\text{, such that } m \text{ is the code of a} \\
& \qquad \text{proof of } P. \\
& \text{if and only if PA} \nvdash P.
\end{aligned}
$$

This equivalence leaves us with two options: *either* $\mathbf{N} \models P$ and PA $\nvdash P$ *or* $\mathbf{N} \not\models P$ and PA $\vdash P$. Notice, however, that the second of these two options contradicts the soundness of PA (since it trivially entails that $\mathbf{N} \not\models$ PA). Hence, we are forced to accept the first option. And this first option looks an awful lot like the claim that P is true but unprovable.

This, therefore, gives us a sketch of the claims Floyd and Putnam want to argue against (that P says "P is unprovable" and that P itself is "true but unprovable"). To understand their objections to these claims, we can begin by looking at their argument for KC.[9] So suppose, just for the sake of argument, that PA is consistent but that we have discovered that PA $\vdash \neg P$. The first thing to notice is that this entails that $\mathbf{N} \not\models$ PA (that is, since $\mathbf{N} \models \mathrm{PA} \Rightarrow \mathbf{N} \models \neg P \Rightarrow \mathbf{N} \not\models P \Rightarrow \mathrm{PA} \vdash P \Rightarrow$ PA is inconsistent). It follows, therefore, that PA is satisfied *only* by nonstandard models of number theory—that is, by models which are not isomorphic to the natural numbers. To use the technical jargon, it follows that PA is *ω-inconsistent*.[10]

---

[9] The argument occurs on pp. 625–27.

Consider, then, a specific model of PA. Call this model **M**, and call the code of P, p_0. Then, since PA $\vdash \neg P$, we know that $\mathbf{M} \models \exists x$ PROOF(x, $\hat{p}_0$). Therefore, there is some element $m \in \mathbf{M}$ such that $\mathbf{M} \models$ PROOF[m, $\hat{p}_0$]. But, by numeralwise representability, $\mathbf{M} \models \neg$PROOF($\hat{n}$, $\hat{p}_0$) for each natural number n. Hence, the relevant m is not one of the ordinary natural numbers—it is, of necessity, one of the "nonstandard" elements of **M**.[11] Further, and this is the second thing to notice here, there is no interesting sense in which this nonstandard m "codes up" a proof of P—or of any other formula, for that matter. Given this, we have no reason to think that the formula PROOF(x, y), *as it gets interpreted by the model* **M**, still captures the notion "y is the code of a sentence and x is the code of a proof of that sentence."

In this situation, therefore, Floyd and Putnam argue that it is hard to justify interpreting P as meaning "P is not provable." After all, the intuitive motivation for interpreting P this way involved interpreting P on the natural numbers—that is, interpreting "+" as plus, "×" as times, and letting "$\exists x$" range over the natural numbers. But now it turns out that this interpretation is incompatible with PA itself (since PA has *only* nonstandard models). Similarly, the interpretation of P as "P is not provable" owed something to the interpretation of PROOF(x, y) suggested by arithmetic expressibility. But, as we have just seen, this interpretation of PROOF(x, y) breaks down when we interpret PROOF(x, y) on nonstandard models. Hence, once we make the assumption that PA $\vdash \neg P$—and thus that *there are no standard models for* PA—we have every reason to give up our initial interpretations of both PROOF(x, y) and of P.[12]

This, then, is the core of Floyd and Putnam's argument for KC. What a given formula "expresses" depends on the model at which we interpret it. If we interpret PROOF(x, y) and P on the natural numbers, then it is plausible to think that they express the facts that "x codes a proof of y" and that "P is unprovable." If we interpret them on nonstandard models, then it is no longer plausible to think that they express such things. Therefore, once we assume that PA is ω-inconsistent—and hence that *all* "admissible interpretations" take place on nonstandard models—we are left with *no* good reasons for interpreting P as "P is not provable."

---

[10] Formally, ω-inconsistency is a bit stronger than the claim that a system has only "nonstandard" models to say that PA is ω-inconsistent means that there is some particular formula, $\phi(x)$, such that for each n, PA $\vdash \neg\phi(\hat{n})$, but it is also the case that PA $\vdash \exists x\, \phi(x)$. In the specific case where PA $\vdash \neg P$, we can actually identify the relevant $\phi(x)$: for each n, PA $\vdash \neg$PROOF($\hat{n}$, $\hat{p}_0$), but PA $\vdash \exists x$ PROOF(x, $\hat{p}_0$). So, in this case, it is the very formula used in the construction of P that serves to witness the ω-inconsistency of PA.

[11] Cf. the discussion of ω-inconsistency in the preceding footnote.

[12] In general, there may also be problems concerning the interpretation of SUBST(x, y, z). In particular, there will be nonstandard elements m_1, m_2, $m_3 \in \mathbf{M}$ such that $\mathbf{M} \models$ SUBST(m_1, m_2, m_3), even though it is clear that m_2 does not code an ordinary natural number and that neither m_1 nor m_3 code formulas. Fortunately, this does not have to affect the interpretation of P, since we *can* show that **M** interprets SUBST($\hat{k}$, $\hat{k}$, z) correctly (that is, $\mathbf{M} \models$ SUBST($\hat{k}$, $\hat{k}$, m) $\Leftrightarrow \mathbf{M} \models m = \hat{p}_0$).

So much for KC itself. With KC in hand, I turn to the consequences Floyd and Putnam draw from KC. To begin, KC shows that the *formal structure* of P does not force us to interpret P as "P is not provable" (since this is an interpretation we would *give up* under certain circumstances—for example, if we discovered that PA $\vdash \neg P$). Nor, Floyd and Putnam argue, do the other details of Gödel's *mathematics* force us to interpret P this way. As formulated in his original paper, Gödel's proof is purely syntactic (using numeralwise representability to show that if PA is ω-consistent, then PA $\nvdash P$).[13] Hence, as far as the *mathematics* goes, we can dispense with *interpretation* altogether.[14] Given all this, it is unclear where the claim that P should be interpreted as meaning "P is unprovable" is supposed to *come from*. At best, this seems to be a "metaphysical claim" (632) which gets grafted onto Gödel's mathematics; it is not, in any interesting sense, something which *follows from* that mathematics.

At the end of the day, Floyd and Putnam take themselves to have vindicated Wittgenstein's skepticism about the common insistence that Gödel's incompleteness theorem shows that there are true but unprovable sentences of ordinary number theory. To make this claim plausible, we seem to need an interpretation of P which makes P *say* "P is not provable." But KC—along with an analysis of the actual *mathematics* of Gödel's proof—undercuts the idea that Gödel's work provides such an interpretation. As a result, Floyd and Putnam conclude that the common insistence that P is "true but unprovable" is more a "metaphysical claim" than a "mathematical result" (632). And, while this does not entail that the claim is *false*, it does show "how little sense we have succeeded in giving it" (632).

---

[13] As Mic Detlefsen has pointed out to me, the claim that Gödel's proof is "purely syntactic" needs to be treated with some care. On the one hand, Gödel's proof does not require specifying a full, formal interpretation of the symbols of his language. (Gödel himself emphasizes this fact on pp. 171, 177, and 181.) Nor does it involve a notion of *truth* for that language. So, it is not semantic in the way that a model-theoretic argument is semantic (or one which made use of arithmetic expressibility).

On the other hand, the argument does involve a systematic *association* between natural numbers and terms in our language (for example, between n and $\bar{n}$). This association is crucial for proving the numeralwise representability results on which the overall proof depends. So, even Gödel's "syntactic proof" involves general correlations between natural numbers and the formal expressions which, in Gödel's own terms, "denote" those numbers.

[14] At best, Floyd and Putnam suggest, Gödel's paper gives rise to a proof-theoretic conception of truth—one under which "P is true" *means* PA $\vdash P$ and "P is false" *means* PA $\vdash \neg P$. Clearly, however, this is an interpretation which leaves no room for "true but unprovable" sentences of arithmetic.

Now, for the reasons mentioned in the last footnote, I find the claim that Gödel's mathematics does not involve "interpretation" somewhat problematic (though there is clearly *a* sense in which it is true) For the sake of argument, though, I will avoid these complexities by simply granting the claim that Gödel's proof is "purely syntactic" and that it involves no "interpretation."

II. ON KC

In this section, I step back to examine Floyd and Putnam's argument for KC in more detail. (In section III, I will look at their more general conclusions concerning the incompleteness theorem.) To see what is wrong with Floyd and Putnam's argument for KC, we need to back up a bit and recall the hypothetical situation KC envisages. Initially, we come to Gödel's theorem with an interpretation of the language of formal arithmetic on the natural numbers—an interpretation which reads "$+$" as plus, "$\times$" as times, and which lets "$\exists x$" range over the natural numbers.[15] Given this interpretation, we proceed to assume two things. First, we assume that our background axiomatization of arithmetic is *sound*—that is, that all of the axioms of PA come out *true* on our interpretation. (If we use standard model-theoretic machinery, this amounts to the claim that $\mathbf{N} \models \mathrm{PA}$.)[16] Second, on the basis of an argument like that given on page 200, we assume that we can interpret P as "P is not provable."

Now at this point, KC suggests the following hypothetical: suppose $\mathrm{PA} \vdash \neg P$. As we saw in section I, this entails that $\mathbf{N} \not\models PA$. Hence, we face a choice. On the one hand, we could modify our background interpretation of arithmetic, giving up $\mathbf{N}$ as an appropriate model for our language and limiting ourselves to those (nonstandard) models which happen to satisfy PA. If we make this choice, then we will be forced to give up the interpretation of P as "P is not provable" (for the reasons highlighted by Floyd and Putnam). On the other hand, we could keep our background *interpretation* of arithmetic language and give up the assumption that PA provides a satisfactory *axiomatization* of arithmetic.

---

[15] After Tarski, this interpretation would probably be fleshed out using model-theoretic machinery. But this is not necessary Gödel himself sketches a non-Tarskian version of the interpretation, and it is this non-Tarskian interpretation which Gödel uses in formulating the notion of arithmetic expressibility (and in proving things about that notion), see pp. 181–91.

[16] For the remainder of the paper, I will assume that we *have* used standard model-theoretic machinery in fleshing out our interpretation. Hence, I will use $\mathbf{N} \models \mathrm{PA}$ to express the soundness of our axioms

It is clear from their paper that Floyd and Putnam think we should take the first option (as witnessed, for instance, by their insistence that only models which satisfy PA should count as "admissible interpretations" for our language).[17] But it is equally clear that they provide no real *argument* for preferring this option. Instead, they simply ignore the possibility of *keeping* $\mathbb{N}$ as the canonical interpretation for our language while *abandoning* (or, at the very least, *modifying*) the ω-inconsistent axiomatization that stands in conflict with this interpretation.[18]

So far, this is simply an objection to Floyd and Putnam's *argument* (pointing out a lacuna which could, in principle, be filled in). But there is a deeper problem here: Floyd and Putnam's assumptions about the (hypothetical) response of the mathematical community to the discovery that PA $\vdash \neg P$ are almost certainly false. Although such a discovery would cause a great deal of consternation, I think the vast majority of mathematicians would look for ways of revising PA in order to block the proof in question—that is, would try to isolate the specific axioms of PA that are essential to the proof and to eliminate some of them from our axiomatization.[19] There are three points to make about this.

First, it is almost *unimaginable* that mathematicians would adopt recognizably nonstandard models of arithmetic as canonical for interpreting the language of number theory. Neither would they accept a provably ω-inconsistent axiomatization of arithmetic as adequate (since it would not, after all, describe the natural numbers!). Given this, the straightforward claim that mathematicians would reject the interpretation of P as "P is not provable" *because* they accept nonstandard models as the basis for interpreting arithmetic is surely mistaken

---

[17] The insistence that we limit ourselves to models of PA when we interpret arithmetic runs rather deep in Floyd and Putnam's paper. At one point they even suggest that, were we to find PA *inconsistent*, we should conclude that there are *no* admissible interpretations of arithmetic (see 626).

[18] This is not, perhaps, so surprising in Putnam's case, as a similar insistence on the priority of axioms over interpretations lies at the heart of his so-called "model-theoretic argument" against realism. See sections III–IV of Timothy Bays, "On Putnam and His Models," this JOURNAL, XCVIII, 7 (July 2001). 331–50.

[19] My guess is that mathematicians would initially focus on the uses of induction in the proof. The hope would be that some well-motivated restriction of the induction scheme would enable us both to restore ω-consistency and to understand *why* our initial scheme went wrong (for example, perhaps we allowed induction on some subtly-paradoxical predicate/formula). This seems *far* more likely than the abandonment of $\mathbb{N}$ which Floyd and Putnam urge upon us.

(though it is equally surely what Floyd and Putnam's argument requires).

Second, although mathematicians would not abandon the interpretation of P as "P is not provable" for the reasons Floyd and Putnam suggest, there is one reason they might abandon it. Suppose it turns out that the specific mathematics used in proving arithmetic expressibility is somehow implicated by the discovery that $\mathrm{PA} \vdash \neg P$. (Say, because there is some $\phi \in \mathrm{PA}$ such that (1) ϕ is crucial to the proof of $\neg P$ and (2) there are numbers n and m such that m codes a formula, and n codes a proof of that formula, but $\mathrm{PA}\setminus\{\phi\} \nvdash \mathrm{PROOF}(\bar{n}, \bar{m})$.)[20] In such a case, the modifications to PA which are needed to *avoid* the acceptance of nonstandard models might undercut various arithmetic expressibility results. In doing so, they might also undercut the traditional interpretation of P.

This, then, looks like a circumstance where something like KC might really be true. Still, there are some obvious worries. First, the supposition in the last paragraph is extremely implausible. The numeralwise representability results on page 199 can be proved in *very* weak fragments of arithmetic, and it is hard to imagine these fragments being problematic.[21] Second, any discovery of problems in weak arithmetic—say, in Q_0—would require such deep revisions of present mathematics that it is virtually impossible to adjudicate questions concerning "what we would/should do" in such circumstances. In particular, neither I nor Floyd and Putnam are in a position to intelligently evaluate KC under this kind of hypothetical. Finally, whatever we would do under this hypothetical, we would do it for reasons other than those Floyd and Putnam suggest. If, for instance, we would abandon the interpretation of P as "P is not provable," then we would do so as part of a larger revision of PA aimed at *keeping* the standard model of arithmetic; we would not accept nonstandard models in order to keep the ω-inconsistent PA.

This brings me to my third point: nothing I have just said implies that mathematicians would *lose interest* in PA and its models (that is, under the circumstances envisioned by KC). Just as logicians now study nonstandard models of arithmetic (while acknowledging that these models *are* nonstandard), so mathematicians would continue to think about nonstandard models after the discovery that PA $\vdash \neg P$ (indeed, they would probably study them a lot more than they do now). Nevertheless, the majority of mathematical work—and the work that would be universally recognized as *arithmetic*—would continue to involve the study of **N**: it would involve interpreting the language of arithmetic *on* **N**, and it would look for axioms satisfied *by* **N**.[22]

---

[20] Of course, this particular result would simply block the proof of arithmetic expressibility *via* numeralwise representability. The crucial case—and the case we are really interested in—is where the elimination of ϕ would undercut *every* proof of arithmetical expressibility (perhaps by undercutting the recursive definition of satisfaction). Nevertheless, the case above illustrates the *kind* of problem at issue here.

[21] So, for instance, numeralwise representability holds for systems as weak as Robinson's Q_0 (a fragment of arithmetic which involves *no* induction); it is very hard to imagine Q_0 being ω-inconsistent.

These, then, are some reasons for thinking that Floyd and Putnam's argument for KC is flawed and that KC itself is mistaken. Before moving on, I think it is useful to digress a moment and examine a claim similar to KC outside the mathematical context. Suppose we have a formal axiomatization of some part of physics—call it T and suppose that it is formulated in the language $\mathcal{L}$. Next, suppose we discover that T proves things that conflict with the basic physical phenomena T was supposed to describe (for example, T makes a series of glaringly false predictions). Then, just as in Floyd and Putnam's case, we face a choice: we can abandon T and start looking for new axioms that better describe the phenomena we are interested in, or we can abandon the physical phenomena and start studying arbitrary models of T (numerical models, perhaps).

Here, I think it is obvious that the physics community would take the first option. T was interesting only (or, at least, primarily) because it seemed to describe a specific class of physical phenomena (that is, described it via the standard interpretation of $\mathcal{L}$). Once T goes wrong about such phenomena, then T has to be modified. In the scientific case, therefore, preserving the original interpretation of our language turns out to be far more important than preserving our original axiomatization. I see no reason to think that things are different for the mathematical case; hence, I see no reason to accept Floyd and Putnam's argument for KC. At the very least, the structural similarities between their argument and the (obviously faulty) physical analogue should lead us to be highly suspicious of the former.

III. ON GÖDEL'S THEOREM

In this section, I turn from Floyd and Putnam's defense of KC to make some brief comments concerning the conclusions they draw from this claim. First, I concede a point: there is nothing in the formal structure of P—that is, in P's very syntax—which forces us to interpret P as "P is not provable." Nor, as Floyd and Putnam notice, does Gödel's original proof require such an interpretation. So, if these were the only ways of providing a mathematically significant interpretation of P, then Floyd and Putnam would be right in challenging the claim that P means "P is not provable."

---

[22] A qualification is in order here. A lot of current work in arithmetic focuses on the purely algebraic properties of the natural numbers (and associated structures like **Z** and **Q**). To the extent that nonstandard models of **N** (and the associated nonstandard versions of **Z** and **Q**) continue to satisfy the relevant algebraic axioms, people working on algebraic number theory may not care about nonstandardness. Nevertheless, no matter how interesting these nonstandard models are (to algebrists and logicians), the canonical core of arithmetic will continue to be the study of **N**, and the axioms for arithmetic will have to be axioms satisfied by **N**.

Fortunately, these are *not* the only respectable ways of interpreting P. The natural interpretation of the language of arithmetic—the interpretation under which "$+$" means plus, "$\times$" means times, and "$\exists x$" ranges over the natural numbers—can be made perfectly rigorous. At present, it would be most natural to do this using model-theoretic machinery or Tarski's original apparatus of language, metalanguage, meta-metalanguage, and so forth. But even without this machinery, we can formulate the interpretation in a mathematically perspicuous manner. Indeed, Gödel himself uses this interpretation in his 1931 paper, both to formulate the notion of arithmetic expressibility and to prove theorems involving this notion.[23]

Further, this interpretation has closer ties to the incompleteness theorem than Floyd and Putnam's argument would lead one to believe. I have already noted that Gödel *uses* the interpretation in several parts of his paper, and he seems to *presuppose* it in many of his informal asides.[24] More substantially, the interpretation helps to explain why Gödel's theorem is *interesting* in the first place. The theorem is not interesting because it shows that a randomly selected axiom system happens to be incomplete: there are *lots* of incomplete axiom systems, and most of them are rather boring.[25] Instead, the theorem shows that a standard axiomatization of *arithmetic*—of the theory of $\mathbf{N}$—is incomplete. More significantly, the theorem does not just show that *this particular* axiomatization is incomplete. Rather, it shows that arithmetic is *intrinsically* incomplete: no recursive extension of PA (or even of Q_0) provides a complete axiomatization of arithmetic.[26] It is this intrinsic incompleteness of *arithmetic* which makes Gödel's first theorem so interesting.

Finally, and most importantly, although it is certainly true that Gödel's original proof of incompleteness was syntactic, there is an alternate, semantic proof which makes essential use of the interpretation at issue here. This proof starts with the arithmetical expressibility results from page 200 and then uses the argument which immediately follows those results (again, page 200) to show that *P* is a true but unprovable sentence (and, hence, that *P* is undecidable from PA).[27] So, even though Gödel did not use this interpretation in his official proof of incompleteness, the interpretation is still very closely related to the underlying *mathematics* of Gödel's paper.[28]

---

[23] See Gödel, p. 181 for the definition of arithmetic expressibility See pp. 183–87 for some theorems which use this notion.

[24] See Gödel, pp. 149–51.

[25] For example, the systems $\emptyset$, $\{P(c)\}$, and $\{c \neq d\}$ are all trivially, and uninterestingly, incomplete.

[26] So, there is no way to "fix" the incompleteness of PA by adding a few new axioms.

When all is said and done, then, I think that Floyd and Putnam are mistaken in the conclusions they draw from KC. There is a perfectly good—and a perfectly *mathematically* respectable—interpretation of the language of arithmetic under which *P* expresses the fact that *P* is not provable. On this interpretation, Gödel's theorem really does show that *P* is "true but not provable." Further, this interpretation is in no way foreign to Gödel's work. Gödel uses the interpretation in the paper where he proves his incompleteness theorem; the interpretation helps to explain the *significance* of that theorem; and the interpretation can be used—as Gödel well knew—to provide a rigorous proof of that theorem. All of this is plain, simple mathematics; there are no "metaphysical claims" anywhere on the horizon.

IV. "RUSSELL'S SYSTEM" AND FOUNDATIONS

Before concluding, I want make two remarks concerning an issue I bypassed near the beginning of this paper: the fact that I have formulated my arguments in terms of PA, while Wittgenstein, Floyd and Putnam formulated theirs in terms of "Russell's system." Now on one level, any reversion back to Russell's system would only make my argument stronger. PA is a far more widely accepted formal system than Russell's system is (or ever was). Hence, just as it is clear that we would modify PA to deal with a discovery that PA is ω-inconsistent (as argued in section II), it is *even more clear* that we would modify Russell's system to deal with an ω-inconsistency in that context.[29] As a result, the arguments I gave in sections II and III would be even stronger if KC were reformulated in terms of Russell's system. So, there is nothing "slippery" in moving the arguments to PA for reasons of perspicuousness.

---

[27] It is worth noting that this semantic proof amounts to a formalization of the informal proof given in the first paragraph of Wittgenstein's remarks (see pages 197–98). Wittgenstein, however, goes on to criticize this proof in his second paragraph. Surprisingly, it is just this criticism which Floyd and Putnam mean to be defending. See Steiner for more on Wittgenstein's rejection of the semantic proof.

[28] Two historical comments are in order. First, Gödel himself was clearly aware of the semantic proof of incompleteness. he sketches it at the beginning of his 1931 paper (pp. 149–51), and he outlines it again in a series of lectures given at Institute for Advanced Study in 1934. Second, Gödel had several reasons for avoiding this semantic proof in his 1931 paper. For one thing, the syntactic proof provides crucial "ingredients" for proving the second incompleteness theorem (while the semantic proof does not). For another, Gödel intended his theorem to have relevance to David Hilbert's program, but only the syntactic proof of the theorem could have such relevance. Hence, Gödel had a number of purely mathematical reasons for using a syntactic proof in 1931. Given this, Gödel's use of the syntactic proof should not be interpreted as a lack of *awareness* of the semantic proof (and still less as a *rejection* of that proof).

[29] Two technical issues deserve comment here. First, the notion of ω-inconsistency depends on the version of Russell's system with which we are working. Godel himself

On another level, though, there may seem to be a problem here. Russell wanted his system to provide a *foundation* for mathematics: to provide the framework in which other parts of mathematics are formulated, and to set the standards for mathematical rigor. Given this, it might seem odd to think that we can "step outside" this framework to discuss its semantics (and then *reject* the framework if it does not live up to our expectations!). To put the point in Floyd and Putnam's terms: "to confess that this is what one has to do would be to abandon the claim for the *foundational* status of a system such as *Principia Mathematica* entirely" (630). On this view, then, it may seem that my entire argument in sections II and III depends on adopting the wrong *attitude* toward our background axiomatization (on treating it as "just one piece of mathematics among others"),[30]

Although this line may initially be tempting, it should be rejected for two reasons. First, there *cannot* be a problem simply with stepping back to study the metatheory of our axiomatization. For, if that were the problem, then it would indict the study of *proof theory* as much as the study of semantics (and, hence, indict the syntactic argument in Gödel's original paper). Any version of Gödel's theorem will involve some amount of "stepping back" from our foundational system (that

---

worked with a version of Russell's logic which was "superimposed" on ordinary Peano Arithmetic: in particular, Gödel's formulation proved that "every object is a number." For this formulation of Russell's system, the definition of ω-inconsistency given in footnote 10 works just fine.

If we work with the full system of PM—or, for that matter, with a system like ZFC—then our system will prove that there are objects other than numbers (and we will have to use a predicate or formula to pick out the particular objects which we want to count as "natural numbers"). For such systems, we say that T is ω-inconsistent if there is a formula, $\phi(x)$, such that for each n, $T \vdash \neg\phi(n)$, but it is also the case that $T \vdash \exists x\,[\mathrm{NUMBER}(x)\ \&\ \phi(x)]$. This is a/the standard way of thinking about ω-inconsistency for systems which talk about more than number theory. (Note that if we simply used the definition from footnote 10, then systems like T would be *trivially* ω-inconsistent.)

Second, the discovery that PM was ω-inconsistent would leave us with several *different* options. We could modify the underlying *logic* of PM; we could change the way we formulate arithmetic *in* PM; or we could leave everything as it is and cease to regard our system as a formulation *of arithmetic* (see the comments in the main text vis-à-vis this last option). The only thing we cannot do, I contend, is to leave things as they are *and* insist that we still have a formulation of *arithmetic*.

[30] For what it is worth, Gödel himself seems to share my "bad attitude" toward Russell's system. For Gödel, Russell's system was simply one of several "related systems" to which his results applied, and he felt free to step back and discuss both the syntax and the semantics of each of these systems.

is why it is a result in *metamathematics*), and I see no principled reasons for permitting us to step back to make syntactic generalizations, while forbidding us to step back to make semantic generalizations.[31] Therefore, if the foundationalist worry is to be cogent, it has to focus on the claim that metatheoretic analysis can make us *give up* a foundational system (and not on the very idea of metatheoretic analysis).

Second, there are clear cases where we *can* be forced to give up a foundational system (for example, when it proves a contradiction).[32] And even when we do not give up a foundational system—because it is too clear, precise and elegant to give up—we can still reject it *as a foundation for a specific discipline*. If a "foundation for arithmetic" winds up proving that $2 + 2 = 5$, then it is not a foundation for arithmetic (though it may well be foundation for something else). Similarly, if a "foundation for analysis" proves that the derivative of $x^3 + 1$ is x^{27} then it is not a foundation for analysis (though, again, it may be a foundation for something else).[33] It is just a *crazy* view of foundational research which holds that, once a foundational system has been proposed, we can no longer legitimately discuss whether that system is *effective* at founding the subjects we wanted it to found in the first place.

In the end, then, I do not see any problems with either my move from Russell's system to PA or my willingness to consider modifications of "foundational" axiom systems. In particular, I do not see any "foundationalist" reasons for changing the conclusions I reached in sections II and III. There is a perfectly good—and, indeed, a perfectly canonical—interpretation of arithmetic under which Wittgenstein's *P* really does say "*P* is not provable." Given this interpretation, Gödel's theorem helps to show that there are "true but unprovable" sentences of ordinary number theory. Nothing in Wittgenstein's remarks—or in Floyd and Putnam's analysis of those remarks—should lead us to think otherwise.

TIMOTHY BAYS

University of Notre Dame

---

[31] As a point *tu quo*, I would note that Floyd and Putnam themselves engage in purely semantic reasoning about Russell's system. So, for instance, they argue that it would be acceptable to define "truth in PM" as "holding in all models of PM" (631). Similarly, they are willing to prove that, if PM $\vdash \neg P$, then *all models of PM* are nonstandard (625). Given this, it would be unreasonable for them to object *in principle* to the semantic analysis of foundational systems.

[32] The important point, here, is that we can recognize from the *outside* that contradictions are unacceptable. We do not just accept contradictions because our "foundational" system tells us to. Hence, we certainly have *some* ability to step back and engage in critical reflection on (purported) foundations.

[33] And, I would argue, if a "foundation for arithmetic" winds up being ω-inconsistent, then it is not really a foundation for arithmetic (whatever other virtues it might happen to have).

BOOK REVIEWS

Reasonably Vicious. CANDACE VOGLER. Cambridge: Harvard University Press, 2002. viii + 295 p. Cloth $47.00.*

Candace Vogler's insightful and provocative book takes on several projects that unfold one from another, as if new philosophical aims were being discovered along the way. Its initial aim is to explain why it is that instrumentalist theories of practical reasoning have such grip on us, such that there seems to be a recurring need to attempt to refute them. In explaining that, Vogler distinguishes between instrumentalism as such, which includes a "bifurcationist" moral psychology that radically differentiates the cognitive from the volitional, and "calculative" theories of practical reasoning, which do not, but instead concentrate on the end-means form of "practical justification." Her explanation of instrumentalism's grip on us is, roughly, that instrumentalism is easily confused with the view that all practical justification is calculative. The second task, then, becomes shoring up the plausibility of the latter thought. The third layer of the project concerns the particular way that she sets out to shore this up, namely by resurrecting Aquinas's division of worldly goods into the useful, the pleasant, and the fitting. Five of the book's eight chapters are devoted to elaborating and expounding this division. Finally, as an apparently independent aim, and seemingly as an afterthought, the book arrives at its title theme, addressing the connection between calculative practical reasoning and morality. Vogler argues that an amoralist does not necessarily suffer from any calculative practical irrationality.

Vogler's initial diagnostic chapter usefully distinguishes among (a) the metaphysics and epistemology of value, (b) the moral psychology of action, (c) the structure of practical reasoning, and (d) the structure of practical justification (12). Her principal point in making these distinctions is to suggest that the opponents of so-called "instrumentalism" have been sloppy about them and have supposed that a thesis in one of these categories (such as the assertion of a Humean, bifurcationist moral psychology) necessarily combines with a thesis in another (such as the assertion that the structure of practical justifi-

---

* This review has benefited immeasurably from extensive discussion of Vogler's book with William Haines, Samuel Kerstein, Mark C Murphy, Nancy Sherman, and Karen Stohr.

0022-362X/04/0104/211-15

cation is necessarily "calculative," or end-means). The first two categories are dropped from consideration early on, and, as the book proceeds, it becomes increasingly clear that Vogler is devoting all of her attention to the last, the "structure of practical justification." How are the final two categories distinguished?

The distinction Vogler is drawing between the structure of practical reasoning and the structure of practical justification becomes fully clear only in chapter 7. There, she addresses those unnamed philosophers who have argued that sound practical reasoning can settle upon new final ends that are not means (causal or constitutive) to ends with which the deliberator started (157). In response, Vogler argues that even assuming this is possible, still, deliberation itself is an action or activity, and as such proceeds with an end in view, namely *figuring out what to do* (166). Here, she might have quoted Kant: "every action has its end."[1] Accordingly, all deliberation or practical reasoning, no matter what its structure or content, is calculative in *form*, as it is directed to figuring out what to do (169).

What follows from this seemingly innocuous observation? Figuring out what Vogler takes to follow from it is complicated by the book's intermixture of diagnosis and more straightforward argument. For many of its pages, it seems she may simply be aiming to give the calculative view a run for its money in order to indicate how instrumentalist views could ever have had such a hold on us, whereas in other places she is clearly defending theses in her own voice. Among the theses she appears to be defending are a weaker and a stronger, each thrice repeated. The weaker thesis amounts to a denial of the kind of reasons externalism long ago defined by the late Bernard Williams:

> Noncalculative theorists insist that there can be reasons for an agent to *B* quite apart from her plans, projects, particular attachments, character, skills, habits, or further ends; the calculative theorist denies this (50, 150, 185).

Thus, at one level, Vogler's defense of the calculative theory of practical reasoning comes down to insisting that practical reasoning can generate reasons for action only by virtue of drawing on ends that antecedently have a foothold in the agent's motivational set. Unlike bifurcationist defenders of internalism, however, Vogler seems happy to let the truth of the calculative theory, so described, rest on the conceptual fact that, because practical deliberation is a kind of action, it necessarily begins with an end in view, namely that of figuring out what to do. If this is all that internalism, or the calculative view, entails, then there is indeed little reason to resist it.

---

[1] *The Metaphysics of Morals* Mary Gregor, trans. (New York: Cambridge, 1996), Ak. 385.

Vogler also provides a stronger statement of the calculative thesis about practical justification, however—one that she is also out to defend:

> Let *A* and *B* be actions of different types. One has reason to *A* only if one takes it that *A*-ing is a means to (or part of) attaining (or making it possible to attain) a further end, *B*-ing, and one wants to *B* (for no particular reason or because *B*-ing is pleasant, useful, or fitting) (48, 73, 149).

Here is a seemingly controversial version of the calculative theory—and also the link to Aquinas's tripartite division of goods. The unnamed philosophers who argue that sound practical reasoning can establish new ends would presumably want to contest this restriction on reasons for action—or would they?

Before we can answer this question, we must attempt to clear up some apparent anomalies in this stronger formulation of the calculative theory. We need to understand (i) why this thrice-repeated formulation would be structured so as apparently to demand a regress of practical reasons, (ii) how it can help make intelligible that one has any practical reasons, and (iii) why it requires that the agent believe there to be an end-means connection between *B* and *A*. The first issue arises because this is plainly put forward as a universal generalization across actions or action-types. Suppose that one has located a pair of actions, *A* and *B*, that one takes to be related as means to end, and that one wants to *B*. Then one has satisfied this necessary condition for having a reason to *A*. But does one have a reason to *B* just by virtue of wanting to (for no particular reason, say)? No, for the thesis again applies to *B*: there must be some other action *C* to which *B* is related as a (causal or constitutive) means if the agent is to have any reason to *B*. And so on *ad infinitum*. Vogler nowhere addresses this regress of reasons as a worry: that is the first puzzle with this stronger formulation. It is intimately connected with the second puzzle: How can this end-means relationship help make a reason for action intelligible if it is enough that *B* is something that the agent wants "for no particular reason"? If this answer is available, why is it not already available at the first layer, enabling the agent to evade this supposed necessary condition simply by saying "I felt like *A*-ing"? Finally, does it not seem crazy to hold that an agent has a reason for acting only if she "takes it" that a relevant end-means connection holds? Consider Williams's classic description of the agent who "believes that this stuff is gin, when it is in fact petrol. Has he reason, or a reason, to mix this stuff with tonic and drink it? On the one hand, it is just very odd to say that he has a reason to drink it.... On the other hand, if he does drink it, we not only have an explanation of his doing so (a reason why he did it), but we have such an explanation which is of the reason-for-action form."[2] Now, Vogler's stronger thesis states only a necessary condition for the existence of a practical reason, not a sufficient one, so she is not committed to saying that Williams's agent has a reason to drink the glass of gas. Conversely, however, she is committed to the claim that if, unbeknownst to the agent, the glass contains, not gin but an antidote to a poison that is about to kill her, still the agent has no reason to drink it unless she wants some gin. This seems equally odd.

Williams's description, however, provides the clue we need to resolve this puzzle as well as the other two. Although Vogler pointedly drops the theology to which each was committed, she follows G.E.M. Anscombe even more closely than she follows Aquinas. In particular, she finds Anscombe's concentration on actual actions, and the question why they were done, to be considerably more fruitful than the "deliberationist" question of how we reason about which actions to take. In Williams's original petrol/gin case, "if he does drink it," and we ask why he did, we *will* want to know what he took there to be in the glass. By the same token, if he did not take there to be any crucially needed antidote in the glass, then there will be no mystery why he refrained from drinking. Following Anscombe, Vogler is concentrating on the justification of actions actually done.

This Anscombian focus on the enterprise of explaining actions already done helps explain Vogler's differentiation, which may have seemed mysterious when presented above, between the structure of practical reasoning and the structure of practical justification. It also helps with the other two puzzles about her stronger formulation. The regress of reasons is not a worry, as there is no ambition to locate the ultimate normative grounds of action. Rather, to "justify" an action in the relevant sense, it appears, is simply to lay out the end-means form without which, as Aristotle and Anscombe have helped us see, no movement can be grasped as a case of action. That this is what Vogler means is confirmed by an elaborate thought experiment in chapter 6, in which she asks us to imagine what the practical world would look like if the "calculative order" were wholly drained out of it. Affirming that "the calculative form revealed by answers to Anscombe's 'Why?' question belongs to intentional action as such," she notes that if the calculative relation were removed from the world, actions would decompose into meaningless fragments (130f.). Thus, to the question "Why are you crossing the road?" you would have available only the answer, "because I felt like it" (or "for no particular reason"). But, she suggests, that is also the only answer one would have to the questions "Why are you stepping off the curb?" and "Why are you lifting your left foot?" Remove the calculative form, and intentional action as such collapses. Conversely, the "justification" with which she is concerned is provided so long as this calculative form, conceptually necessary to all intentional action as such, is restored. That is why a regress vainly in search of a stopping point to "Why?" questions is not a concern and why intelligibility is sufficiently provided, for Vogler's purposes, by indicating an action's calculative form.

---

[2] "Internal and External Reasons," in *Moral Luck* (New York: Cambridge, 1981), pp. 101–13, here see p. 102.

Accordingly, Vogler's stronger formulation, on this reading, is not nearly as controversial as it had appeared. The unnamed philosophers who defend the possibility of reasoning practically about final ends can accept even this stronger formulation, on the understanding that it is simply formulating a necessary condition on an agent's reasons in acting (as she actually is or has), not a necessary condition on the reasons that bear on an agent's deciding how to act. By the same token, it should be easy to accept her title claim: vicious actions can enjoy the "reasonableness" of the calculative form. In other words, evil action is possible.

With this interpretation taken as read, Vogler's explanation of why instrumentalism has had such a hold on us makes perfect sense: it has had a firm hold on us because all intentional action is calculative in form. As Kant said, "every action has its end." Setting aside, with Vogler, issues about the moral psychology of practical reasoning, what we can say is that instrumentalist theories of practical reasoning have retained a grip on us because of our persistent tendency to confuse the form of intentional action with the structure of practical reasoning. Calling this kind of calculative view a theory of practical "justification" is somewhat misleading; but as long as we pay sufficient attention to Vogler's frequent invocations of Anscombe and to her disinterest in "deliberative" accounts of practical reasoning, we will be able to remember that she is really talking about the reasoned structure of intentional action as such.

In thus attempting to clarify the core elements of Vogler's argument, I have not had space to describe her fascinating explorations of the pleasant, the useful, and the fitting. Philosophical readers are bound to find reading these chapters pleasant, useful, or a fitting use of their time.

HENRY S. RICHARDSON

Georgetown University

NEW BOOKS

SOMERVILLE, JAMES: *The Epistemological Significance of the Interrogative.* Burlington, VT: Ashgate, 2002, xi, 272 p. Cloth $79.95.

SOULE, EDWARD: *Morality and Markets: The Ethics of Government Regulations.* Lanham, MD: Rowman and Littlefield, 2003. xiv, 193 p. Cloth $70.00, paper $23.95.

SPADER, PETER H.: *Scheler's Ethical Personalism: Its Logic, Development, and Promise.* New York: Fordham, 2002. xviii, 327 p. Cloth $45.00, paper $20.00.

SPELMAN, ELIZABETH V.: *Repair: The Impulse to Restore in a Fragile World.* Boston: Beacon, 2002. 165 p. Cloth $24.00.

STAMOS, DAVID N.: *The Species Problem: Biological Species, Ontology, and the Metaphysics of Biology.* Lanham, MD: Lexington, 2003. 381 p. Cloth $80.00.

STANISTREET, PAUL: *Hume's Scepticism and the Science of Human Nature.* Burlington, VT: Ashgate, 2002. xi, 226 p. Cloth $69.95.

STECKER, ROBERT: *Interpretation and Construction: Art, Speech, and the Law.* Malden, MA: Blackwell, 2003. x, 212 p. Paper $27.95.

STERELNY, KIM: *Thought in a Hostile World: The Evolution of Human Cognition.* Malden, MA: Blackwell, 2003. xi, 262 p. Cloth $35.00.

STOKHOF, MARTIN: *World and Life as One: Ethics and Ontology in Wittgenstein's Early Thought.* Stanford: University Press, 2002. xvi, 329 p. Cloth $60.00.

STONEHAM, TOM: *Berkeley's World: An Examination of the Three Dialogues.* New York: Oxford, 2002. xviii, 308 p. Cloth $65.00, paper $21.95.

STREVENS, MICHAEL: *Bigger than Chaos: Understanding Complexity through Probability.* Cambridge: Harvard, 2003. xii, 413 p. Cloth $59.95.

SULLIVAN, STEFAN: *Marx for a Post-Communist Era: On Poverty, Corruption, and Banality.* New York: Routledge, 2002. x, 191 p. Cloth $80.00, paper $24.95

SWANTON, CHRISTINE: *Virtue Ethics: A Pluralistic View.* New York: Oxford, 2003. xi, 312 p. Cloth $35.00.

SWINBURNE, RICHARD: *The Resurrection of God Incarnate.* New York: Oxford, 2003. vii, 224 p. Cloth $70.00, paper $24.95.

TANTILLO, ASTRIDA ORLE: *The Will to Create. Goethe's Philosophy of Nature.* Pittsburgh: University Press, 2002. xii, 241 p. Cloth $32.50.

TAYLOR, CRAIG: *Sympathy: A Philosophical Analysis.* New York: Palgrave Macmillan, 2002. ix, 155 p. Cloth $55.00.

TAYLOR, RICHARD: *Virtue Ethics: An Introduction.* Amherst, NY: Prometheus, 2002. xvi, 125 p. Paper $18.00.

TEMES, PETER S.: *The Just War: An American Reflection on the Morality of War in Our Time.* Chicago: Ivan R. Dee, 2003. xv, 219 p. Cloth $25.00.

THORNHILL, CHRIS: *Karl Jaspers: Politics and Metaphysics.* New York: Routledge, 2002. x, 272 p. Cloth $95.00.

TIMMONS, MARK: *Moral Theory: An Introduction.* Lanham, MD: Rowman and Littlefield, 2002. x, 291 p. Cloth $65.00, paper $18.95.

TOULMIN, STEPHEN E.: *The Uses of Argument,* Updated Version. New York: Cambridge, 2003. xiv, 247 p. Cloth $55.00, paper $20.00.

TUCKNESS, ALEX: *Locke and the Legislative Point of View: Toleration, Contested Principles, and the Law.* Princeton: University Press, 2002. xiii, 206 p. Cloth $49.50, paper $17.95.

TURNER, STEPHEN P.: *Brains/Practices/Relativism.* Chicago: University Press, 2002. ix, 214 p. Paper $19.00.

UMPHREY, STEWART: *Complexity and Analysis.* Lanham, MD: Lexington, 2002. vii, 351 p. Cloth $90.00, paper $26.95.

UNGER, PETER: *Philosophical Relativity.* New York: Oxford, 2003. x, 132 p. Paper $19.95.

URBINATI, NADIA: *Mill on Democracy from the Athenian Polis to Representative Government.* Chicago: University Press, 2002. xii, 293 p. Cloth $37.50.

VALLICELLA, WILLIAM F.: *A Paradigm Theory of Existence: Onto-theology Vindicated.* Boston: Kluwer, 2002. xii, 281 p. Cloth $91.00.

VAN FRAASSEN, BAS C.: *The Empirical Stance.* New Haven, CT: Yale, 2002. xix, 282 p. Cloth $30.00.

VAN INWAGEN, PETER: *Metaphysics,* Second Edition. Boulder, CO: Westview, 2002. xi, 235 p. Cloth $82.00, paper $26.00.

VAUGHN, LEWIS, and AUSTIN DACEY: *The Case for Humanism: An Introduction.* Lanham, MD: Rowman and Littlefield, 2003. ix, 223 p. Cloth $75.00, paper $26.95.

VELKLEY, RICHARD L.: *Being after Rousseau: Philosophy and Culture in Question.* Chicago: University Press, 2002. x, 192 p. Cloth $40.00, paper $18.00.

VERBEEK, BRUNO: *Instrumental Rationality and Moral Philosophy: An Essay on the Virtues of Cooperation.* Boston: Kluwer, 2002. xii, 286 p. Cloth $82.00.

VERENE, DONALD PHILLIP: *The Art of Humane Education.* Ithaca, NY: Cornell, 2002. xiii, 80 p. Cloth $19.95; *Knowledge of Things Human and Divine: Vico's New Science and Finnegans Wake.* New Haven, CT: Yale, 2003. xiv, 264 p. Cloth $45.00.

VOGLER, CANDACE: *Reasonably Vicious.* Cambridge: Harvard, 2002. viii, 295 p. Cloth $45.00.

WALDRON, JEREMY: *God, Locke, and Equality: Christian Foundations of John Locke's Political Thought.* New York: Cambridge, 2002. xii, 263 p. Cloth $60.00, paper $22.00.

WALTON, DOUGLAS: *Legal Argumentation and Evidence.* University Park: Penn State UP, 2002. xvii, 374 p. Cloth $65.00; *Relevance in Argumentation.* Mahwah, NJ: Erlbaum, 2004. xvi, 311 p. Cloth $69.95, paper $34.50.

WARNOCK, MARY: *Making Babies: Is There a Right Way to Have Children?* New York: Oxford, 2002. vi, 120 p. Cloth $13.95, paper $12.95.

WARREN, JAMES: *Epicurus and Democritean Ethics: An Archaeology of Ataraxia.* New York: Cambridge, 2002. xiv, 241 p. Cloth $55.00.

WEATHERSTON, MARTIN: *Heidegger's Interpretation of Kant: Categories, Imagination and Temporality.* New York: Palgrave Macmillan, 2002. xi, 209 p. Cloth $62.00.

WEGNER, DANIEL M.: *The Illusion of Conscious Will.* Cambridge: MIT, 2002. xi, 405 p. Cloth $34.95.

WEHANE, PATRICIA H., and TARA J. RADIN, with NORMAN E. BOWIE: *Employment and Employee Rights.* Malden, MA: Blackwell, 2003. x, 228 p. Paper $29.95.

WEISS, BERNHARD: *Michael Dummett.* Princeton: University Press, 2002. ix, 197 p. Cloth $55.00, paper $18.95.

WEISS, PAUL: *Surrogates.* Bloomington: Indiana UP, 2002. xviii, 181 p. Cloth $34.95.

WEISSMAN, DAVID: *Lost Souls: The Philosophic Origins of a Cultural Dilemma.* Albany: SUNY, 2003. xiii, 210 p. Cloth $57.50, paper $18.95.

WEITHMAN, PAUL J.: *Religion and the Obligations of Citizenship.* New York: Cambridge, 2002. xi, 227 p. Cloth $55.00.

WESTPHAL, KENNETH R.: *Hegel's Epistemology: A Philosophical Introduction to the Phenomenology of Spirit.* Indianapolis: Hackett, 2003. xvi, 146 p. Cloth $34.95, paper $12.95.

WILLIAMS, BERNARD: *Truth and Truthfulness: An Essay in Genealogy.* Princeton: University Press, 2002. xi, 328 p. Cloth $27.95.

WINCHESTER, JAMES J.: *Aesthetics across the Color Line: Why Nietzsche (Sometimes) Can't Sing the Blues.* Lanham, MD: Rowman and Littlefield, 2002. xi, 145 p. Cloth $60.00, paper $19.95.

WINGO, AJUME H.: *Veil Politics in Liberal Democratic States.* New York: Cambridge, 2003. xvii, 162 p. Cloth $55.00, paper $22.00.

WIRTH, JASON M.: *The Conspiracy of Life: Meditations on Schelling and His Time.* Albany: SUNY, 2003. ix, 290 p. Cloth $71.50, paper $23.95.

WISNOVSKY, ROBERT: *Avicenna's Metaphysics in Context.* Ithaca, NY: Cornell, 2003. viii, 305 p. Cloth $65.00.

WITT, CHARLOTTE: *Ways of Being: Potentiality and Actuality in Aristotle's Metaphysics.* Ithaca, NY: Cornell, 2003. x, 161 p. Cloth $35.00.

WOLFF, JONATHAN: *Why Read Marx Today?* New York: Oxford, 2002. viii, 136 p. Cloth $20.00.

WOODS, JOHN: *Paradox and Paraconsistency: Conflict Resolution in the Abstract Sciences.* New York: Cambridge, 2003. xviii, 362 p. Cloth $70.00, paper $26.00.

WRIGHT, CRISPIN: *Saving the Differences: Essays on Themes from Truth and Objectivity.* Cambridge: Harvard, 2003. ix, 549 p. Cloth $55.00.

YOUNG-AH GOTTLIEB, SUSANNAH: *Regions of Sorrow: Anxiety and Messianism in Hannah Arendt and W.H. Auden.* Stanford: University Press, 2003. xii, 306 p. Cloth $65.00, paper $24.95.

YURACKO, KIMBERLY A.: *Perfectionism and Contemporary Feminist Values.* Bloomington: Indiana UP. x, 170 p. Cloth $44.95, paper $19.95.

YUSA, MICHIKO: *Zen and Philosophy: An Intellectual Biography of Nishida Kitaro.* Honolulu: Hawaii UP, 2002. xxvi, 482 p. Paper $29.95.

ZAHAVI, DAN: *Husserl's Phenomenology.* Stanford: University Press, 2003. viii, 182 p. Cloth $45.00, paper $15.95.

ZIPORYN, BROOK: *The Penumbra Unbound: The Neo-Taoist Philosophy of Guo Xiang.* Albany: SUNY, 2003. ix, 186 p. Cloth $68.50, paper $22.95.

ŽIŽEK, SLAVOJ: *For They Know Not What They Do: Enjoyment as a Political Factor,* Second Edition. New York: Verso, 2002. cvii, 288 p. Paper $21.00; *The Puppet and the Dwarf: The Perverse Core of Christianity.* Cambridge: MIT, 2003. viii, 188 p Paper $16.95.

ZUCKERT, MICHAEL P.: *Launching Liberalism: On Lockean Political Philosophy.* Lawrence: Kansas UP, 2002. xi, 375 p. Cloth $40.00, paper $19.95.

ZUPKO, JACK: *John Buridan: Portrait of a Fourteenth-century Arts Master.* Notre Dame: University Press, 2003. xix, 446 p. Cloth $70.00, paper $40.00.

NEW BOOKS: PAPERBACK REPRINTS

ARON, RAYMOND: *The Dawn of Universal History: Selected Essays from a Witness of the Twentieth Century.* YAIR REINER, ed., BARBARA BRAY, trans. New York: Basic Books, 2002/2003. xxvi, 518 p. Paper $22.00.

BECKER, CARL L.: *The Heavenly City of the Eighteenth-Century Philosophers, Second Edition.* New Haven: Yale, 1932/2003. xix, 168 p. Paper $16.00.

BILLINGTON, RAY: *Living Philosophy: An Introduction to Moral Thought,* Third Edition. New York: Routledge, 1988/2003. xviii, 354 p. Paper $27.95.

BROUGHTON, JANET: *Descartes's Method of Doubt.* Princeton: University Press, 2002/2003. xv, 217 p. Paper $16.95.

BROWN, JAMES ROBERT: *Who Rules in Science?* Cambridge: Harvard, 2001/2004. xiii, 236 p. Paper $16.95.

CHATTERJEA, TARA: *Knowledge and Freedom in Indian Philosophy.* Lanham, MD: Lexington, 2002/2003. xvi, 161 p. Paper $24.95.

DAVIDSON, ARNOLD I.: *The Emergence of Sexuality: Historical Epistemology and the Formation of Concepts.* Cambridge: Harvard, 2001/2004. xvi, 254 p. Paper $19.95.

GEUSS, RAYMOND: *Public Goods, Private Goods.* Princeton: University Press, 2001/2003. xxvi, 148 p. Paper $14.95.

HADOT, PIERRE: *What Is Ancient Philosophy?* MICHAEL CHASE, trans. Cambridge: Harvard, 2002/2004. xii, 362 p. Paper $15.95.

HAKSAR, VINIT: *Rights, Communities, and Disobedience: Liberalism and Gandhi.* New York: Oxford, 2001/2003. xxxi, 202 p. Paper $14.95.

KANT, IMMANUEL: *Observations on the Feeling of the Beautiful and Sublime.* JOHN T. GOLDTHWAIT, trans. Berkeley: California UP, 1991/2003. 124 p. Paper $16.95.

LACOCQUE, ANDRÉ, and PAUL RICOEUR: *Thinking Biblically: Exegetical and Hermeneutical Studies.* DAVID PELLAUER, trans. Chicago: University Press, 1998/2003. xix, 441 p. Paper $18.00.

LAMPERT, LAURENCE: *Nietzsche's Task: An Interpretation of Beyond Good and Evil.* New Haven: Yale, 2001/2004. x, 320 p. Paper $24.00.

LOVIBOND, SABINA: *Ethical Formation.* Cambridge: Harvard, 2002/2004. xv, 203 p. Paper $19.95.

MARGALIT, AVISHAI: *The Ethics of Memory.* Cambridge: Harvard, 2002/2004. xi, 227 p. Paper $14.95.

O'HAGAN, TIMOTHY: *Rousseau.* New York: Routledge, 1999/2003. xvi, 320 p. Paper, $27.95.

PUTNAM, HILARY: *The Collapse of the Fact/Value Dichotomy and Other Essays.* Cambridge: Harvard, 2002/2004. x, 190 p. Paper $16.95.

RICHARDSON, HENRY S.: *Democratic Autonomy: Public Reasoning about the Ends of Policy.* New York: Oxford, 2002/2004. xii, 316 p. Paper $19.95.

RICOEUR, PAUL: *The Just.* DAVID PELLAUER, trans. Chicago: University Press, 2000/2003. xxiv, 161 p. Paper $14.00.

SCHEFFLER, SAMUEL: *Boundaries and Allegiances: Problems of Justice and Responsibility in Liberal Thought.* New York: Oxford, 2001/2002. 221 p. Paper $19 95.

SCRUTON, ROGER: *A Short History of Modern Philosophy: From Descartes to*

Wittgenstein, Second Edition. New York: Routledge, 1995/2002. ix, 318 p. Paper $14.95.

SELIGMAN, ADAM B.: *Modernity's Wager: Authority, the Self, and Transcendence.* Princeton: University Press, 2000/2003. xii, 177 p. Paper $19.95.

SEN, AMARTYA: *Rationality and Freedom.* Cambridge: Harvard, 2004. ix, 736 p. Paper $19.95.

SHAND, JOHN: *Philosophy and Philosophers: An Introduction to Western Philosophy.* Montreal: McGill-Queen's, 1993/2002. x, 341 p. Paper $19.95.

SMART, BARRY: *Michel Foucault, Revised Edition.* New York: Routledge, 1985/2002. xv, 150 p. Paper $14.95.

STEINER, MARK: *The Applicability of Mathematics as a Philosophical Problem.* Cambridge: Harvard, 1998/2002. viii, 215 p. Paper $19.95.

TAMEN, MIGUEL: *Friends of Interpretable Objects.* Cambridge: Harvard, 2001/2004. 194 p. Paper $16.95.

TAYLOR, CHARLES: *Varieties of Religion Today: William James Revisited.* Cambridge: Harvard, 2002/2003. vii, 127 p. Paper $12.00.

TAYLOR, MARK C.: *The Moment of Complexity: Emerging Network Culture.* Chicago: University Press, 2001/2003. xii, 340 p. Paper $17.00.

VAN DEN BERG, AXEL: *The Immanent Utopia: From Marxism on the State to the State of Marxism.* New Brunswick, NJ: Transaction, 1989/2003. xli, 580 p. Paper $29.95.

WELTON, DONN: *The Other Husserl: The Horizons of Transcendental Phenomenology.* Bloomington: Indiana UP, 2000/2002. xvi, 496 p. Paper $29.95.

WIERENGA, EDWARD R.: *The Nature of God: An Inquiry into Divine Attributes.* Ithaca, NY: Cornell, 1989/2003. xii, 238 p. Paper $24.95.

WILLIAMS, BERNARD: *Truth and Truthfulness: An Essay in Genealogy.* Princeton: University Press, 2002/2004. xi, 328 p. Paper $17.95.

WOODRUFF, PAUL: *Reverence: Renewing a Forgotten Virtue.* New York: Oxford, 2001/2002. x, 248 p. Paper $11.95.

THE JOURNAL OF PHILOSOPHY

Celebrates 100 Years

Special Limited Time Offer

Full Volumes 1-99

1904-2002

$50.00 per Volume

All volumes are available for $50.00 plus $5.00 shipping and handling.

Send check or money order to:

The Journal of Philosophy
1150 Amsterdam Avenue, MC 4972
Columbia University
New York, NY 10027
(212) 666-4419
www.journalofphilosophy.org

ADORNO ANSCOMBE AYER BACHELARD
[illegible]
CASSIRER [illegible] CHOMSKY
CHURCH CHURCHLAND DAVIDSON
DENNETT DERRIDA DEWEY DUCASSE
DUMMETT DWORKIN EVANS
FEYERABEND [illegible] FRANKENA
FRANKFURT [illegible] FRIEDMAN
GOODMAN [illegible] HEMPEL
HINTIKKA [illegible] KATZ KIM
KRIPKE KUHN LANGER LEWIS
MACINTYRE MALCOLM MARCUS MARCUSE
MCDOWELL MEAD NAGEL NUSSBAUM
NOZICK PARFIT PEIRCE PLANTINGA
PUTNAM QUINE RAWLS REICHENBACH
RICOEUR RORTY ROYCE RUSSELL
SANTAYANA SARTRE SCANLON SEARLE
SELLARS [illegible]
SWINBURNE [illegible] VAN
FRAASSEN [illegible] WILLIAMS
WITTGENSTEIN [illegible]

CAMBRIDGE

KNOW THEIR THOUGHT

Charles Taylor
Edited by Ruth Abbey
Contemporary Philosophy in Focus
$60.00 Hardback 0-521-80136-2 232pp
$20.00 Paperback 0-521-80522-8

Practical Conflicts
New Philosophical Essays
Edited by Peter Baumann and Monika Betzler
$70.00 Hardback 0-521-81271-2 352pp
$26.00 Paperback 0-521-01210-4

Ancient Anger
Perspectives from Homer to Galen
Edited by Susanna Morton Braund and Glenn W. Most
Yale Classical Studies 32
$65.00 Hardback 0-521-82625-X 336pp

The Cambridge Companion to Abelard
Edited by Jeffrey E. Brower and Kevin Guilfoy
Cambridge Companions to Philosophy
$65.00 Hardback 0-521-77247-8 382pp
$25.00 Paperback 0-521-77596-5

The Cambridge Companion to Thomas Reid
Edited by Terence Cuneo and René van Woudenberg
Cambridge Companions to Philosophy
$70.00 Hardback 0-521-81270-4 392pp
$26.00 Paperback 0-521-01208-2

Kant on Representation and Objectivity
A.B. Dickerson
$60.00 Hardback 0-521-83121-0 228pp

An Introduction to Rights
William A. Edmundson
Cambridge Introductions to Philosophy and Law
$55.00 Hardback 0-521-80398-5 240pp
$22.00 Paperback 0-521-00870-0

Self-Interest before Adam Smith
A Genealogy of Economic Science
Pierre Force
Ideas in Context 68
$65.00 Hardback 0-521-83060-5 296pp

Word and World
Practice and the Foundations of Language
Patricia Hanna and Bernard Harrison
$85.00 Hardback 0-521-82287-4 432pp
$30.00 Paperback 0-521-53744-4

Kant: *Prolegomena* to Any Future Metaphysics
With Selections from the *Critique of Pure Reason*
Second Edition
Edited by Gary Hatfield
Cambridge Texts in the History of Philosophy
$50.00 Hardback 0-521-82824-4 270pp
$18.00 Paperback 0-521-53535-2

Philosophy and the Emotions
Edited by Anthony Hatzimoysis
Royal Institute of Philosophy Supplements 52
$26.00 Paperback 0-521-53734-7 260pp

The Cambridge Companion to Brentano
Edited by Dale Jacquette
Cambridge Companions to Philosophy
$65.00 Hardback 0-521-80980-0 344pp
$24.00 Paperback 0-521-00765-8

www.cambridge.org

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE

WITH CAMBRIDGE

Metaphor and Emotion
Language, Culture, and Body in Human Feeling
Zoltán Kövecses
Studies in Emotion and Social Interaction
$25.00 Paperback 0-521-54146-8 240pp

The End of Art
Donald Kuspit
$28.00 Hardback 0-521-83252-7 224pp

Locke's Philosophy of Language
Walter R. Ott
$55.00 Hardback 0-521-83119-9 168pp

The Paradoxes of Art
A Phenomenological Investigation
Alan Paskow
$65.00 Hardback 0-521-82833-3 272pp

Literal Meaning
François Recanati
$55.00 Hardback 0-521-79246-0 188pp
$20.00 Paperback 0-521-53736-3

Can a Darwinian be a Christian?
The Relationship Between Science and Religion
Michael Ruse
$18.00* Paperback 0-521-63716-3
c 256pp
Now in paperback

Money and the Early Greek Mind
Homer, Philosophy, Tragedy
Richard Seaford
$80.00 Hardback 0-521-83228-4 382pp
$29.00 Paperback 0-521-53992-7

Jung and the Making of Modern Psychology
The Dream of a Science
Sonu Shamdasani
$28.00 Paperback 0-521-53909-9 404pp

States and Citizens
History, Theory, Prospects
Edited by Quentin Skinner and Bo Strath
$60.00 Hardback 0-521-83156-3 244pp
$22.00 Paperback 0-521-53926-9

The Evidence for the Top Quark
Objectivity and Bias in Collaborative Experimentation
Kent W. Staley
$70.00 Hardback 0-521-82710-8 360pp

An Introduction to Mill's Utilitarian Ethics
Henry R. West
$55.00 Hardback 0-521-82832-5 228pp
$20.00 Paperback 0-521-53541-7

Descartes' *Meditations*
An Introduction
Catherine Wilson
Cambridge Introductions to Key Philosophical Texts
$55.00 Hardback 0-521-80981-9 284pp
$20.00 Paperback 0-521-00766-6

Lectures on Logic
Edited by J. Michael Young
The Cambridge Edition of the Works of Immanuel Kant
$40.00* Paperback 0-521-54691-5 c 727pp

Prices subject to change

www.cambridge.org CAMBRIDGE UNIVERSITY PRESS

HEAVEN'S FRACTAL NET

Retrieving Lost Visions in the Humanities

William J. Jackson

Explores the beautiful and intriguing fractal patterns found in art, literature, religion, and beyond.

cloth $59.95 paper $29.95

JOHN DEWEY AND MORAL IMAGINATION

Pragmatism in Ethics

Steven Fesmire

A broad pragmatist grounding that revitalizes ethics and moral philosophy.

cloth $49.95 paper $19.95

INDIANA UNIVERSITY PRESS

at bookstores • order: 1-800-842-6796 • iupress.indiana.edu

Studies in Continental Thought

John Sallis, Editor

THINKING THROUGH FRENCH PHILOSOPHY

The Being of the Question

Leonard Lawlor

A powerful and original engagement with France's most influential philosophers

cloth $59.95 paper $24.95

RACE AND RACISM IN CONTINENTAL PHILOSOPHY

Edited by Robert Bernasconi with Sybol Cook

New thinking on how Continental philosophers approach race and racism.

cloth $59.95 paper $24.95

TRUTH AND GENESIS

Philosophy as Differential Ontology

Miguel de Beistegui

Develops a new philosophy that is grounded in both artistic and scientific truth

cloth $60.00 paper $29.95

THINKING WITH HEIDEGGER

Displacements

Miguel de Beistegui

An invitation to think with Heidegger and draw his ideas out into a more contemporary idiom.

cloth $59.95 paper $24.95

HEIDEGGER'S *CONTRIBUTIONS TO PHILOSOPHY*

An Introduction

Daniela Vallega-Neu

Short, clear, readable introduction to Heidegger's second most important work.

cloth $39.95 paper $19.95

RETHINKING NATURE

Essays in Environmental Philosophy

Edited by Bruce V. Foltz and Robert Frodeman

A landmark volume for defining the terrain of environmental philosophy.

cloth $59.95 paper $24.95

SCHELLING NOW

Contemporary Readings

Edited by Jason M. Wirth

Places Schelling at the core of contemporary philosophical debates.

cloth $49.95 paper $22.95

THE NEW KIERKEGAARD

Edited by Elsebet Jegstrup

New essays on Kierkegaard that read his philosophy in the light of deconstruction.

cloth $59.95 paper $24.95

THE NEW HUSSERL

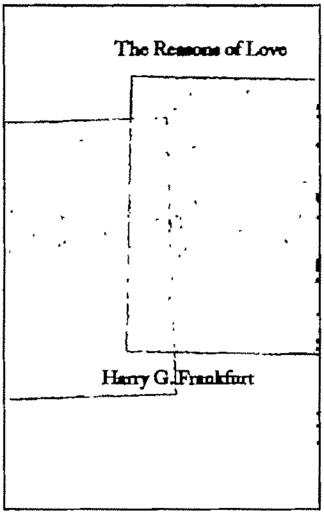
A Critical Reader

Edited by Donn Welton

Presents fresh research and new perspectives on Husserl's thought

cloth $75.00 paper $29.95

Wise.



The Reasons of Love

Harry G. Frankfurt

This beautifully written book by one of the world's leading moral philosophers argues that the key to a fulfilled life is to pursue wholeheartedly what one cares about, that love is the most authoritative form of caring, and that the purest form of love is, in a complicated way, self-love.

According to Harry Frankfurt, the most basic and essential question for a person to raise about the conduct of his or her life is not what he or she should care about but what, in fact, he or she cannot help caring about

Cloth $19.95 ISBN 0-691-09164-1 Due May

New in paperback

Truth and Truthfulness

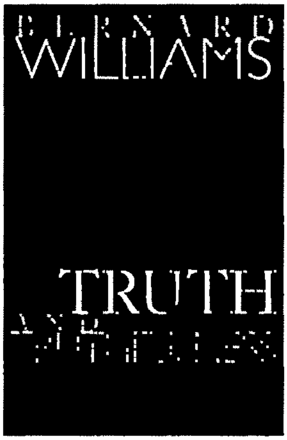
An Essay in Genealogy

Bernard Williams

"This vigorous, crystalline book is an intellectual landmark." —Richard Sennett, *Times Literary Supplement*

"Anyone who wants to understand the relations between the relatively arcane issues concerning truth debated by philosophy professors, and the larger question of what self-image we human beings should have, would do well to read Williams's new book. It is a major work." —Richard Rorty, *London Review of Books*

Paper $17 95 ISBN 0-691-11791-8 Due March



PRINCETON
University Press

800-777-4726 • READ EXCERPTS ONLINE
WWW.PUP PRINCETON EDU

Philosophy from Yale

Karl Jaspers: A Biography

Navigations in Truth

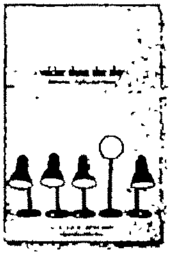
Suzanne Kirkbright

This insightful biography is the first to draw on the full range of Jaspers' diaries and correspondence, and it illuminates not only the German philosopher's life and relationships but also the ideas he developed in his works. $35.00

Wider Than the Sky

The Phenomenal Gift of Consciousness

Gerald M. Edelman, M.D.



In this book, **Nobel Prize winner Gerald M. Edelman "argues persuasively and with exceptional clarity** that consciousness is a process and gives a superb account of its brain mechanisms. A distillation of knowledge and wisdom, this book is a rare, polished gem."
—Apostolos Georgopoulos $24.00

Decades of Dissent

Fifty Years of Political and Social Criticism

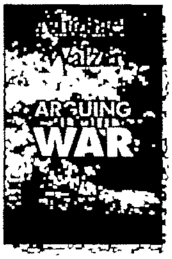
Edited by Nicolaus Mills and Michael Walzer

With an introduction by Mitchell Cohen

"I find this volume of **essays impressive** not only in their quality but also in their surprising relevance to political life today."—Robert Dahl
$25.00

Arguing About War

Michael Walzer



"Walzer's **discussions about war and ethics are always provocative, well argued, and insightful.** He is able to transform complex issues into readable, understandable, and persuasive prose."
—Jean Bethke Elshtain $25.00

The Jewish Political Tradition

Volume two: Membership

Edited by Michael Walzer, Menachem Lorberbaum, and Noam J. Zohar

Co-edited by Ari Ackerman

"[This] second volume is as comprehensive as the first, and one looks forward to the completion of the series, which will provide **an indispensable guide to the evolution of Jewish political thought.**"—*Choice* $40.00

The Jewish Political Tradition

Volume 1: Authority

Edited by Michael Walzer, Menachem Lorberbaum, and Noam J. Zohar

Co-editor Yair Lorberbaum

This book, that launches a landmark four-volume collaborative work, "**is one of the most ambitious Jewish intellectual efforts of recent years.**"
—David Novak, *New Republic*

Selected as a finalist for the 2000 National Jewish Book Awards in the Thought category $25.00

Yale University Press [illegible]

Paperbacks from Yale

The Enlightenment and the Intellectual Foundations of Modern Culture

Louis Dupré

"**This immensely readable book** will cause readers to rethink the Enlightenment and to see its positive aspects. It will also add crucial historical perspective to current discussions of modernity."—Donald Verene $45.00

Confessio Philosophi and Papers Concerning the Problem of Evil, 1671–1678

G. W. Leibniz

Translated and with an introduction and notes by Robert C. Sleigh, Jr.

With contributions from Brandon Look and James Stam

This volume contains papers that represent **Liebniz's early thoughts** on the problem of evil, centering on a dialogue, the *Confessio philosophi*, in which he formulates a general account of God's relation to sin and evil that becomes a fixture in his thinking. The Yale Leibniz Series $70.00

Justice in Plainclothes

A Theory of American Constitutional Practice

Lawrence G. Sager

"This book is **elegantly written, subtly argued,** and full of interesting asides and original observations."
—Michael Seldman $40.00

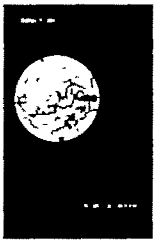
New in paper

One World

The Ethics of Globalization; Second Edition

Peter Singer

With a new preface by the author



"**Timely and thoughtful**. . . . [There is] a refreshing intellectual integrity in Singer's efforts to assess the facts on the ground."—Andrés Martinez, *New York Times Book Review*
$14.00 A Yale Nota Bene paperback

Kant's Transcendental Idealism

An Interpretation and Defense; Revised and Enlarged Edition

Henry E. Allison

"Those seeking a readable, readily available, and reliably literal translation of Kant's brief but influential work cannot do better than this new edition. . . . **This is the translation I would want my own students to have in hand.**"—Susan Shell, *Claremont Review of Books*
$25.00 paperback

The Empirical Stance

Bas C. van Fraassen

"**An ambitious, absorbing, iconoclastic**, metaphilosophical treatise."
—Richard Rorty, *Notre Dame Philosophical Reviews* $20.00 paperback

Nietzsche's Task

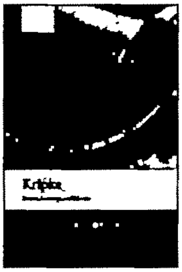
An Interpretation of
Beyond Good and Evil

Laurence Lampert

"**A masterpiece of insightful commentary** and a brilliant interpretation of one of Nietzsche's richest—and most poorly understood—books."
—Graham Parkes $24.00 paperback

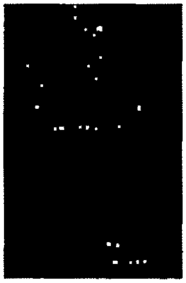
Yale University Press [illegible]

New from
Oxford

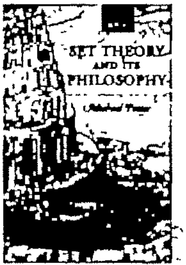


Kripke
Names, Necessity, and Identity
CHRISTOPHER HUGHES
Saul Kripke, in a series of classic writings of the 1960s and 1970s, changed the face of metaphysics and philosophy of language. Christopher Hughes offers a careful exposition and critical analysis of Kripke's central ideas about names, necessity, and identity, and in the process makes significant contributions to continuing debates about such topics as modality, essence, natural kinds, and the relation between the mental and the physical.
April 2004 $49 95

New in Paperback!
Truth and the End of Inquiry
A Peircean Account of Truth
C J. MISAK
"The exposition and defense of Pierce's views on truth is clear, well organized, and largely nontechnical. It interacts with a good deal of the recent literature relevant to pragmatism and truth."
—*Choice*
paper April 2004 $26 95



Studies in the Philosophy of Logic and Knowledge
Edited by THOMAS BALDWIN AND TIMOTHY SMILEY
Questions about knowledge, and about the relation between logic and language, are at the heart of philosophy. Eleven distinguished philosophers from Britain and America contribute papers on such questions. The first half of the book constitutes a running debate about knowledge, evidence and doubt. The second half tackles questions about logic and its relation to language.
2004 $55 00



Set Theory and Its Philosophy
An Introduction
MICHAEL POTTER
Anyone wishing to work on the logical foundations of mathematics must understand set theory, which lies at its heart. What makes the book unique is that it interweaves a careful presentation of the technical material with a penetrating philosophical critique. Potter does not merely expound the theory dogmatically but at every stage discusses in detail the reasons that can be offered for believing it to be true.
April 2004 paper $29 95 cloth $74 00

The Divine Lawmaker
Lectures on Induction, Laws of Nature, and the Existence of God
JOHN FOSTER
John Foster begins by developing a solution to the problem of induction—a solution that involves the postulation of laws of nature, as forms of natural necessity. He then offers a radically new account of the nature of such laws and the distinctive kind of necessity they involve. Finally, he uses this account as the basis for an argument for the existence of God as the creator of the laws and the universe they govern
2004 $39 95

Prices are subject to change and apply only in the US. To order, or for more information, please call 1-800-451-7556. Visit our website at www.oup.com/us

OXFORD

Calcutta University Library
ALIPORE CAMPUS

THE JOURNAL OF PHILOSOPHY

VOLUME CI, NUMBER 5
MAY 2004

221 *Likelihood, Model Selection, and the Duhem-Quine Problem* Elliott Sober

COMMENTS AND CRITICISM
242 *Evidence, Explanation, and Experience: On the Harder Problem of Consciousness* Jakob Hohwy

REVIEW ESSAYS
255 *Inclusive Rationality:* A Review of Amartya Sen: *Rationality and Freedom* Isaac Levi

Published by The Journal of Philosophy, Inc.

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Akeel Bilgrami, Arthur C. Danto, Kent Greenawalt, Patricia Kitcher, Philip Kitcher, Isaac Levi, Mary Mothersill, Philip Pettit, Carol Rovane, Achille C. Varzi. *Editor Emeritus*: Sidney Morgenbesser. *Consulting Editors*: David Albert, John Collins, James T. Higginbotham, Charles D. Parsons, Wilfried Sieg. *Managing Editor*: John Smylie.

THE JOURNAL OF PHILOSOPHY is owned and published by the Journal of Philosophy, Inc. *President*, Arthur C. Danto; *Vice President*, Akeel Bilgrami; *Secretary*, Daniel Shapiro; *Treasurer*, Barbara Gimbel; *Other Trustees*: Lee Bollinger, Leigh S. Cauman, Kent Greenawalt, Michael J. Mooney, Lynn Nesbit.

All communications to the Editors and Trustees and all manuscripts may be sent to John Smylie, Managing Editor, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, New York 10027 FAX: (212) 932-3721. You may also visit our website at: www.journalofphilosophy.org

THE JOURNAL OF PHILOSOPHY

2004

SUBSCRIPTIONS (12 issues)

Individuals	$35.00
Libraries and Institutions	$75.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$15.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are required.

Published monthly as of January 1977; typeset and printed by Capital City Press, Montpelier, VT.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL's terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

POSTMASTER: Periodical postage paid at New York, NY, and other mailing offices.

POSTMASTER: Send address changes to the *Journal of Philosophy* at MC 4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027

© Copyright 2004 by the Journal of Philosophy, Inc. ISSN 0022-362X

THE JOURNAL OF PHILOSOPHY
VOLUME CI, NO. 5, MAY 2004

LIKELIHOOD, MODEL SELECTION, AND THE DUHEM-QUINE PROBLEM*

The Duhem-Quine problem is usually formulated *deductively* with a choice described *dichotomously*. When the conjunction of a hypothesis (H) and an auxiliary assumption (A) entails an observational prediction (O) that fails to come true, should one reject H or reject A? A more general formulation would be to ask what one should say when the conjunction (H & A) confers some probability on O, and instead of considering the two choices just mentioned, the problem would be to evaluate judgments that are a matter of degree. For example, if the observational outcome disconfirms the conjunction (H & A), what determines whether and how much each conjunct is disconfirmed? Indeed, the negative cast of this question can be discarded by generalizing further: How does the disconfirmation or confirmation of the conjunction affect the disconfirmation or confirmation of the conjuncts?[1] The epistemological holism associated with Pierre Duhem[2] and W.V. Quine[3] denies that evidence bearing on (H & A) can have an impact on H that differs from the impact it has on A. This holism can take two forms. *Nondistributive holism* asserts that only whole conjunctions are confirmed and disconfirmed, never their constituent conjuncts; *distributive holism* concedes that evidence

---

* My thanks to Julian Barbour, Martin Barrett, Richard Creath, John Earman, Ellery Eells, Branden Fitelson, Malcolm Forster, Michael Friedman, Alan Hájek, Daniel Hausman, Margaret Moore, John Norton, Michael Stölzner, Peter Turney, and the editors of this JOURNAL for useful suggestions.

[1] If the original Duhem-Quine problem concerns a question about acceptance/rejection, how is the problem described here concerning confirmation/disconfirmation related to that original problem? If decisions about acceptance and rejection need to include an evaluation of the evidence at hand, the second problem is part of the first.

[2] *The Aim and Structure of Physical Theory* (Princeton: University Press, 1954).

[3] "Two Dogmas of Empiricism," in *From a Logical Point of View* (Cambridge: Harvard, 1953), pp. 20–46, and *Philosophy of Logic* (Englewood Cliffs, NJ: Prentice-Hall, 1970).

0022-362X/04/0105/221–41

bearing on the conjunction can have an impact on a conjunct, but insists that the effect on one conjunct must be the same as the effect on the other.[4] Holists grant that hypotheses and auxiliary assumptions are often treated differently when predictions fail, but claim that it is *nonevidential considerations*, such as simplicity or conservatism, that does the work.[5] To refute holism, the challenge is to show how *the evidence* can have an effect on hypotheses that differs from its effect on auxiliary assumptions.

Previous attempts to bring probabilistic tools to bear on the Duhem-Quine problem have mainly been Bayesian.[6] It is intrinsic to this approach that one must discuss $\Pr(H|A)$, $\Pr(A|H)$, and the probability of the observations conditional not just on (H & A) but on (H & not$A$) and on (not$H$ & A).[7] It is not essential to assign point values

---

[4] Distributive holists may assert that the effects on H and A are *qualitatively* the same (that is, that both are confirmed or both are disconfirmed) or, more ambitiously, that the effects are *quantitatively* the same (that is, that the degree of confirmation of H is identical with the degree of confirmation of A). For more on this taxonomy of holisms, see my "Quine's Two Dogmas," *Proceedings of the Aristotelian Society*, LXXIV (2000): 237–80.

[5] I do not concede that simplicity is always an extra-evidential consideration, the point here is that this is what holists happen to believe. For discussion, see my *Reconstructing the Past: Parsimony, Evolution, and Inference* (Cambridge: MIT, 1988), and my "Instrumentalism, Parsimony, and the Akaike Framework," *Philosophy of Science*, LXIX (2002): S112–S123

[6] See, for example, J. Dorling, "Bayesian Personalism, the Methodology of Scientific Research Programs, and Duhem's Problem," *Studies in the History and Philosophy of Science*, X (1979): 177–87, Colin Howson and Peter Urbach, *Scientific Reasoning: The Bayesian Approach*, (La Salle, IL: Open Court, 1989), John Earman, *Bayes or Bust?* (Cambridge: MIT, 1992), and Michael Strevens, "A Bayesian Treatment of Auxiliary Hypotheses," *British Journal for the Philosophy of Science*, LII (2001): 515–37. For a nonBayesian treatment, see Deborah Mayo's *Error and the Growth of Experimental Knowledge* (Chicago: University Press, 1996), which analyzes the Duhem-Quine problem within the context of frequentist statistics.

[7] As a simple example of a Bayesian analysis, let us define the degree of confirmation that X receives from Y, $c[X,Y]$, as the ratio $\Pr(X|Y)/\Pr(X)$. So defined, $c[X,Y] > 1$ when Y positively confirms X and $c[X,Y] < 1$ when Y disconfirms X. By Bayes's theorem, this ratio equals $\Pr(Y|X)/\Pr(Y)$; thus, $c[H, \text{not}O] > c[A, \text{not}O]$ if and only if $\Pr(\text{not}O|H) > \Pr(\text{not}O|A)$, which expands to

$$\Pr(\text{not}O\,|H\,\&\,A)\Pr(A\,|H) + \Pr(\text{not}O|\,H\,\&\,\text{not}A)\Pr(\text{not}A\,|H) >$$
$$\Pr(\text{not}O\,|H\,\&\,A)\Pr(H\,|A) + \Pr(\text{not}O|\,\text{not}H\,\&\,A)\Pr(\text{not}H\,|A).$$

Note the occurrence of $\Pr(A|H)$ and of $\Pr(H|A)$ in this expression. If we assume that H and A are probabilistically independent, the inequality reduces to

$$\Pr(\text{not}O\,|H\,\&\,A)\Pr(A) + \Pr(\text{not}O|\,H\,\&\,\text{not}A)\Pr(\text{not}A) >$$
$$\Pr(\text{not}O\,|H\,\&\,A)\Pr(H) + \Pr(\text{not}O|\,\text{not}H\,\&\,A)\Pr(\text{not}H).$$

Notice the prior probabilities of A and of H. I do not mean to beg questions here about the proper definition of degree of confirmation, on which see Branden Fitelson, "The Plurality of Bayesian Measures of Confirmation and the Problem of Measure Sensitivity," *Philosophy of Science*, LXVI (1999): S362–S378. The point is just to identify the kinds of quantities that a Bayesian analysis must evaluate.

to these quantities; the formal treatments require only that value ranges or inequalities among these quantities be provided. The problem with this approach is that these quantities are often difficult to interpret *objectively*. What is the probability of Newton's theory if there are seven planets? What is the probability of there being seven planets, if Newton's theory is true? And what is the probability that the orbit of Uranus will have a certain shape, if Newton's theory is false and there are seven planets? It does no good to treat these probabilities as subjective degrees of belief. This is unsatisfactory because a subjective interpretation has the consequence that one's "solution" to the problem lacks normative force—one has offered no reason to think that disconfirmation *should* be assigned more to one conjunct than to the other.[8] In other words, Bayesianism in the context of the Duhem-Quine problem encounters the same limitations that Bayesianism often confronts in other settings.

In what follows I will discuss an example of the Duhem-Quine problem in which $\Pr(H|A)$, $\Pr(A|H)$, and $\Pr(O|\pm H \,\&\, \pm A)$ (where H is the hypothesis, A the auxiliary assumptions, and O the observational prediction) can be construed objectively; however, only some of those quantities are relevant to the analysis that I provide. The example involves medical diagnosis. The goal is to test the hypothesis that someone has tuberculosis; the auxiliary assumptions describe the error characteristics of the test procedure. Although it can make sense to talk about the objective probability that someone (randomly drawn from a given population) has tuberculosis and it also can make sense to talk about the objective probability that a test procedure has a certain set of error characteristics, neither of these quantities will enter into the analysis. The analysis proceeds entirely via *likelihoods*; what one needs to consider is just the probability of the observations conditional on four conjunctions of the form $(\pm H \,\&\, \pm A)$.[9] It is a special feature of the example that all four of these conjunctions are *simple statistical hypotheses* in the technical sense that each unambigu-

---

Another Bayesian approach would be to compare, not the *change* in probability that O induces in H with the change it induces in A, but the *absolute values* of $\Pr(H|O)$ and $\Pr(A|O)$. Bayes's Theorem entails that $\Pr(H|O) > \Pr(A|O)$ if and only if $\Pr(O|H)\Pr(H) > \Pr(O|A)\Pr(A)$. Prior probabilities occur in this expression, and $\Pr(H|A)$ and $\Pr(A|H)$ are involved as well, since $\Pr(O|H) = \Pr(O|H \,\&\, A)\Pr(A|H) + \Pr(O|H \,\&\, \mathrm{not}A)\Pr(\mathrm{not}A|H)$ and $\Pr(O|A) = \Pr(O|H \,\&\, A)\Pr(H|A) + \Pr(O|\mathrm{not}H \,\&\, A)\Pr(\mathrm{not}H|A)$.

[8] See Earman (*op. cit.*).

[9] In what follows, I use "likelihood" and "likely" in this technical sense—the likelihood of $(H \,\&\, A)$ with respect to the observation O, is $\Pr(O|H \,\&\, A)$, not $\Pr(H \,\&\, A|O)$.

ously confers a probability on the observations.[10] After describing how the likelihood concept applies to the example concerning medical diagnosis, I will show how similar patterns can arise in the context of a second inferential framework—that of H. Akaike's criterion of model selection;[11] this time the example will involve phylogenetic inference.

I. AN EXAMPLE AND ITS LIKELIHOOD ANALYSIS

Suppose you want to find out whether someone, whom I will call "Newman," has tuberculosis. To do this, you need to use a tuberculosis test. But what makes something a tuberculosis test? The basic idea is that tuberculosis tests are input-output devices. A sample drawn from the subject (for example, of blood or saliva) is put into a black box, which then goes into one of two states, which we call "positive" and "negative." A good tuberculosis test has small error probabilities. Errors, of course, come in two forms—the test can produce false positives and false negatives. Table 1 represents the four relevant conditional probabilities that characterize a tuberculosis test; each has the form Pr(± test outcome | ± tuberculosis). It is important to understand these so-called error probabilities in the right way. For example, e_1 is not the probability that someone with a negative test outcome has tuberculosis; rather, it is the probability that someone with tuberculosis will have a negative test outcome. The relevant probabilities are, so to speak, world-to-device, not device-to-world.[12] It might be less misleading to say that these probabilities measure the test's *sensitivity*; they do not tell you how probable it is that a positive or a negative result is mistaken.[13]

---

[10] Although Mayo (*op. cit.*) is no friend of likelihood, her point that a good test of a hypothesis *H* may fail to be a good test of the auxiliary assumption *A* is congenial to the likelihood analysis I will describe.

[11] "Information Theory as an Extension of the Maximum Likelihood Principle," in B. Petrov and F. Csaki, eds., *Second International Symposium on Information Theory* (Budapest: Akademiai Kiado, 1973), pp. 267–81.

[12] Here I adopt terminology from Hartry Field, who talks about world-to-head and head-to-world reliability. See his "'Narrow' Aspects of Intentionality and the Information-theoretic Approach to Content," in Enrique Villanueva, ed., *Information, Semantics, and Epistemology* (New York: Blackwell, 1990), pp. 102–16.

[13] If our interest in using tuberculosis tests is to find out whether someone probably has tuberculosis, why are error characteristics defined in terms of probabilities of the form Pr(± test outcome | ± tuberculosis), rather than in terms of probabilities of the form Pr(± tuberculosis | ± test outcome)? The reason is that the latter quantities depend on how rare or common tuberculosis is, but the former do not. It is a curious fact about our universe that probabilities of the form Pr(effect|cause) are often time-translationally invariant, whereas probabilities of the form Pr(cause| effect) rarely are. For discussion, see my "Temporally Oriented Laws," *Synthese*, xciv (1993): 171–89

TABLE 1

	S has tuberculosis	*S* does not have tuberculosis
+ test result	1-e_1	e_2
− test result	e_1	1-e_2

How would you go about determining what the error characteristics are of a tuberculosis test? One obvious procedure is to assemble people whom you know have tuberculosis and people whom you know do not, and give each person the test. Suppose that a company in Madison, which is interested in developing a test kit for tuberculosis, does this, with 1000 people in each group. The data the company obtains are given in Table 2. How would you use these frequency data to infer the relevant error probabilities? The standard procedure is to use ***maximum likelihood estimation***; you find the estimate that maximizes the probability of the observations. The estimated probabilities for the Madison test procedure are therefore:

(Madison) Pr(+ test result | *S* has tuberculosis) = 997/1000
Pr(+ test result | *S* has no tuberculosis) =2/1000.

Given these probabilities, the *Law of Likelihood* describes how a person's test result should be interpreted. If the test outcome is positive, this result *favors* the hypothesis (H_1) that he or she has tuberculosis over the hypothesis (H_2) that he or she does not.[14] A negative test result has the opposite evidential meaning. The strength of the differential support that a test outcome provides is usually measured by the ***likelihood ratio***, which has a value of 997/2 favoring H_1 if the result is positive and a value of 998/3 favoring H_2 if it is negative.[15] These hefty values indicate that the Madison test can be said to be a *good* test for tuberculosis. The test provides a lot of information, regardless of the outcome.

TABLE 2

MADISON	1000 with Tuberculosis	1000 with no Tuberculosis
+ test result	997	2
− test result	3	998

---

[14] See Ian Hacking, ***The Logic of Statistical Inference*** (New York: Cambridge, 1965); Anthony Edwards, ***Likelihood*** (New York: Cambridge, 1972); and Richard Royall, ***Statistical Evidence: A Likelihood Paradigm*** (Boca Raton, FL: Chapman and Hall, 1997).

[15] Here and in what follows I rely on likelihood ratios to measure strength of evidence. The warning in Fitelson (*op. cit.*) that epistemological conclusions may be sensitive to choice of measure is relevant here.

In addition to the Madison company I have just described, suppose there is a company in Middleton that has been involved in the same project. They also want to develop a tuberculosis test kit, so they also try out their procedure on 1000 people whom they know have tuberculosis and 1000 people whom they know do not. The data they obtain on their kit are given in Table 3. They then use maximum likelihood estimation to estimate the error probabilities of their test:

(Middleton) $\Pr(+\text{ test result} \mid S \text{ has tuberculosis}) = 990/1000$
$\Pr(+\text{ test result} \mid S \text{ has no tuberculosis}) = 5/1000.$

Notice that the Middleton device is inferred to have slightly larger error probabilities, both positive and negative, than the Madison test. Still, it is a pretty good test. If someone has a positive result on the Middleton test, the likelihood ratio of the two hypotheses H_1 (S has tuberculosis) and H_2 (S does not have tuberculosis) is 990/5 favoring H_1, and a negative outcome engenders a likelihood ratio of 995/10 favoring H_2.

TABLE 3

MIDDLETON	1000 with Tuberculosis	1000 with no Tuberculosis
+ test result	990	5
− test result	10	995

We now can return to the original problem of finding out whether Newman has tuberculosis. I introduced the two tuberculosis tests to give this problem a Duhemian twist. Duhem emphasized that physical theories do not entail observational predictions all by themselves, but do so only when conjoined with auxiliary assumptions. Duhem's insight is preserved in the example at hand, even though the relationships are probabilistic, not deductive. Suppose we give Newman a tuberculosis test and obtain a positive result. The probability of obtaining that result depends both on whether Newman has tuberculosis and on whether we used a test kit from Madison or one from Middleton. The four probabilities are represented in Table 4.

TABLE 4

		Possible Auxiliary Assumptions		
		A_1. Madison		A_2: Middleton
	H_1. Newman has tuberculosis	997/1000	>	990/1000
Hypotheses		∨		∨
	H_2: Newman does not have tuberculosis	2/1000	<	5/1000

Notice that there is a *qualitative asymmetry* between what the observational outcome says about the hypotheses H_1 and H_2 and what it says about the auxiliary assumptions A_1 and A_2. Newman's positive test result renders H_1 more likely than H_2, regardless of whether A_1 or A_2 is true. However, whether A_1 is more likely than A_2 depends on which of the hypotheses is true, and this, I am assuming, is something we do not already know. Of course, if we *do* already know whether Newman has tuberculosis, then the observed test result does provide information about whether the test kit came from Madison or from Middleton. However, the information provided is exceedingly modest. If Newman has tuberculosis, the positive test outcome slightly favors A_1 over A_2; the likelihood ratio here is only 997/990. Similarly, if Newman does not have tuberculosis, then the positive result favors A_2 over A_1, with a likelihood ratio of 5/2. On the other hand, if we not only do not know whether Newman has tuberculosis but cannot even assign a probability to this being the case, the test result tells us nothing about the provenance of the test kit.

If the data and the ensuing maximum likelihood estimates of error probabilities from either Madison or Middleton had been different, it could easily have turned out that there is no *qualitative* asymmetry between the observation's impact on the hypotheses and its impact on the auxiliary assumptions. That is, it is possible for the observation to provide information about both. However, this does not mean that the *amount* of information provided must be the same; there still can be a *quantitative* asymmetry, even if there is no qualitative asymmetry. By changing the probability in the lower-right cell in Table 4, we obtain Table 5. Now Newman's positive test result favors H_1 over H_2, regardless of which auxiliary assumption is true, and it also favors A_1 over A_2, regardless of which hypothesis is true. However, the observation provides much more information about whether Newman has tuberculosis than it does about whether the test kit came from Madison or Middleton. With respect to the hypotheses, the ratio is either 997/2 or 990/1, depending on which test procedure was used. With respect to the auxiliary assumptions, the ratio is either 997/990 or 2/1, depending on whether Newman has tuberculosis. Note that 997/2 and 990/1 are both much bigger than 997/990 and 2/1. In this example, the observation provides more information about the hypotheses than it does about the auxiliary assumptions. Of course, the reverse situation can also obtain and it also is possible for the situation to be perfectly symmetrical; two examples of symmetry will be discussed at the end of the paper.

TABLE 5

		Possible Auxiliary Assumptions		
		A_1. Madison		A_2. Middleton
Hypotheses	H_1: Newman has tuberculosis	997/1000	>	990/1000
		∨		∨
	H_2: Newman does not have tuberculosis	2/1000	>	1/1000

When a conjunction (H & A) makes a prediction that neither conjunct makes on its own, epistemological holism says that it is *never possible* for the outcome to have an evidential significance for H that differs from the significance it has for A. The generality of this thesis means that just one counterexample is enough to refute it. I claim that the example just described performs that function. Let H be the hypothesis that Newman does not have tuberculosis and let A be the hypothesis that one is using the Madison test procedure. The conjunction (H & A) predicts that the test result will be negative in the sense that it confers on that outcome a probability of 998/1000. But suppose the test comes out positive. To see what this outcome means for H and what it means for A, we need to know what the alternatives are to each. With the alternatives as described, the test result can have a bearing on the hypotheses that differs fundamentally from the bearing it has on the auxiliary assumptions. Both qualitative and quantitative asymmetries are possible. Epistemological holism is false.

II. TWEAKING THE EXAMPLE

The example just described, in which you do not know beforehand whether Newman has tuberculosis and also do not know which tuberculosis test kit you are using, is somewhat artificial. Scientists typically know the provenance of the test kits they use as well as their estimated error probabilities. But even in this more realistic setting, there still is room for Duhemian puzzlement. Suppose we *know* we are using the Madison test kit. However, we recognize that the error probabilities associated with this test kit are merely estimates—we have no certainty that the estimated values are exactly right. Again we give the test to Newman and again obtain a positive result. Does that outcome provide information about whether Newman has tuberculosis and does it also provide information about the test procedure's error characteristics? If so, does the outcome provide more information about one of these than it does about the other?

It might seem intuitive to say that Newman's test outcome provides zero information about the error characteristics of the test procedure. After all, Newman is quite unlike the 2000 subjects who were used to calibrate the test; we have no independent knowledge as to whether he has the disease. Of course, if we knew that he *probably* has the disease, or that he *probably* does not, that would tell us whether his positive test result is *probably* a true positive or is *probably* a false positive, and that would lead us to modify slightly our estimates of the test's error characteristics. But suppose we do not know even that. How, then, can the test result provide any information at all about the test's error characteristics?

Newman's test outcome could be a false positive or it could be a true positive. Let us consider these possibilities in turn. If Newman's test result is a true positive, we should add this result to the 2000 individuals already studied and change our estimate of the test's error characteristics to

Pr(+ test result | S has tuberculosis) = 998/1001
Pr(+ test result | S has no tuberculosis) =2/1000.

On the other hand, if Newman's test result is a false positive, we should revise our estimate of the test's error characteristics as follows:

Pr(+ test result | S has tuberculosis) = 997/1000
Pr(+ test result | S has no tuberculosis) =3/1001.

Of course, we do not know whether Newman's result is a false positive or a true positive, so we do not know which pair of estimates we should use to characterize the procedure's error characteristics. However, this uncertainty does not prevent us from formulating a pair of *conditional estimates*:

(New Madison) If Newman has tuberculosis, then Pr(+ test result | S has tuberculosis) = 998/1001 and Pr(+ test result | S has no tuberculosis) = 2/1000.

If Newman does not have tuberculosis, then Pr(+ test result | S has tuberculosis) = 997/1000 and Pr(+ test result | S has no tuberculosis) = 3/1001.

We now can ask whether there is a difference in likelihood between the old estimates (Madison) or the new, conditional, estimates (New Madison) that were obtained by taking account of Newman's positive test result.

Table 6 summarizes the situation; cell entries represent the probability of Newman's positive test result, conditional on different combinations of hypotheses and auxiliary assumptions. Notice first that there is a *qualitative symmetry* between what the observation says about the hypotheses and what it says about the auxiliary assumptions. Newman's positive test result renders H_1 more likely than H_2, regardless of which auxiliary assumption is true, and the result also favors (New

Madison) over (Madison) regardless of which hypothesis is true.[16] However, there is a *quantitative asymmetry*. The observation is *very* informative about whether Newman has tuberculosis; 997/2 and 998/3 are both large. In contrast, the observation is only modestly informative about the choice between the auxiliary assumptions; the ratios are approximately 998/997 and 3/2, and these are both rather small. Why does Newman's test result have such a negligible impact on the estimates of the test's error characteristics? The reason is that Newman is just one person out of 2001. Had we initially estimated the error probabilities by using just 200 subjects, or 20, or 2, Newman would have mattered more.

TABLE 6

		Possible Auxiliary Assumptions		
		Madison		New Madison
Hypotheses	H_1: Newman has tuberculosis	997/1000	<	998/1001
		∨		∨
	H_2: Newman does not have tuberculosis	2/1000	<	3/1001

It may seem odd that I even consider (New Madison). If estimates of the error characteristics of a test procedure must be based solely on frequency data, then speculations about what our maximum likelihood estimates would be if we knew whether Newman has tuberculosis are irrelevant. This sensible attitude flies in the face of epistemological holism—it entails that Newman's test outcome provides considerable evidence about whether he has tuberculosis and zero information about the error characteristics of the test procedure. If this were correct, there would be a qualitative as well as a quantitative asymmetry. The analysis in which (Madison) and (New Madison) are compared comes close to this result, but does not coincide with it exactly. I argued that the evidence *slightly* favors (New Madison) over (Madison), not that the observation is literally *informationless*. In terms of the larger picture of seeing what is wrong with epistemological holism, this difference does not matter. But in terms of the specifics of likelihood reasoning, it does.

III. SIGNIFICANCE OF THE TWEAKED EXAMPLE

This last example illustrates a very general fact about the calibration of measurement instruments and the validation of test procedures in science. The typical situation is that the error characteristics of a test procedure are first ascertained and then the procedure is applied to new individuals. One usually does not already know whether these new individuals have the condition being tested (otherwise, why apply the test?); indeed, one often does not even know whether the new individuals *probably* have the condition. Many of us have opinions about the approximate frequency of tuberculosis in this or that population; if we were prepared to assume that Newman was drawn at random from such a population, we would be entitled to talk about the prior probability that he has tuberculosis. Many scientific tests are not like this. Galileo gauged the reliability of his telescope by training it on various *terrestrial* objects. He used it to identify the flags on ships coming over the horizon and the inscriptions on distant buildings; in all these cases it was possible to determine *independently* whether the reports were correct.[17] Galileo then looked through his telescope at *Jupiter*; his observations provided strong evidence that Jupiter has moons, but little or no information about the telescope's error characteristics. Understanding this asymmetry does not require that one assign a prior probability to Jupiter's having moons. This is fortunate, since Galileo was in no position to assign an objective prior probability to that proposition.[18]

---

[16] (Madison) has a lower likelihood than (New Madison) in each row, since $a/b < (a+1)/(b+1)$, if $0<a<b$.

There is a sense in which the likelihood analysis of the tuberculosis example is nonBayesian, but this is not because likelihood is an idea that Bayesianism abhors. On the contrary—likelihood is a fundamental quantity in Bayes's theorem. What I mean is that the analysis does not use the full-blown resources that Bayesianism assumes are available. First, prior and posterior probabilities play no role. Second, a likelihoodist will be happy to compare the likelihoods of two simple statistical hypotheses (S_1 and S_2), but often is loath to compare the likelihoods of a simple hypothesis (S_1) and its negation (notS_1) when that negation is composite. Suppose notS_1 is equivalent to a disjunction of simple hypotheses (S_2 or S_3 or...or S_n). If so, the likelihood of notS_1 will be a *weighted average* of the likelihoods of S_2, S_3,...S_n where the weighting term has the form $\Pr(S_i \mid \text{not}S_1)$. This weighting term often lacks an objective interpretation. If Newton's theory is false, what is the probability of each of the theory's specific alternatives? Thus the problem with priors often recurs as a problem for likelihoods.[19] It is a very special property of the tuberculosis example that the two hypotheses considered are *exhaustive* (assuming that Newman exists) and the four conjunctions are simple.[20]

---

[17] See Phillip Kitcher's *The Advancement of Science: Science without Legend, Objectivity without Illusions* (New York: Oxford, 1993), pp. 228–33, and his "Real Realism: The Galilean Strategy," *Philosophical Review*, cx (2001): 151–98.

[18] Galileo estimated the error characteristics of his telescope by using it in problems that involved relatively small terrestrial distances; he then applied this detection device to an astronomical object that was much farther away. There certainly was room to wonder, at the time, how trustworthy this bold extrapolation was. My point here is not to comment on the legitimacy of Galileo's inference, but to note how often scientists use the protocol I described in connection with the tuberculosis test. The question of how the behavior of the measuring device should be *parameterized* (for example, a single set of error characteristics for sightings of all objects, or two such sets—one for objects that are near and another for objects that are far away), as opposed to the question of how values for parameters should be *estimated*, will be discussed later.

The relations of qualitative and quantitative asymmetry that I have described are purely formal, and therefore do not depend on one's interpretation of probability. Still, the question may be asked of what interpretation of probability I am using when I say that the likelihoods evaluated in the example concerning Newman's tuberculosis are "objective." Clearly, I cannot think of probabilities as subjective degrees of belief. But neither do I wish to endorse the objective interpretations—actual relative frequencies, hypothetical relative frequencies, propensities—now on offer. My preference is the *no-theory theory of probability*, which rejects the need for a reductive analysis of what probability statements mean. Probability is a theoretical quantity. It obeys the axioms of probability and it bears nondeductive inferential relations to observed relative frequencies. Probability, like other theoretical magnitudes, cannot be reduced to observations, nor does it need to be.[21]

This brings us to the question of how general the treatment provided here of Newman's tuberculosis test can be said to be. Does the structure of this quotidian example apply to all situations in which the Duhem-Quine problem arises? There are two types of situation in which it does not; I will describe one of them now and postpone the other until the next section. Scientists sometimes react to the predictive failure of a conjunction by formulating an alternative to one conjunct without bothering to formulate an alternative to the other. For example, when John Crouch Adams and Jean Joseph Le Verrier tried to account for the fact that the conjunction of Newton's theory and the assumption that there are seven planets generates an inaccurate prediction of Uranus's orbit, they did not try to invent an alternative to Newton's theory.[22] Rather, they set their minds to constructing a specific alternative to the auxiliary assumption, and the result was the prediction and confirmation of a new planet, which we now call Neptune.[23] In this case, the comparison was between two conjunctions—(Newton & seven planets) and (Newton & eight planets)—first using the old observations of Uranus's orbit, and then assembling new observations drawn from pointing telescopes in the right direction.[24] This one-sided response to observational anomaly also occurs when the auxiliary assumptions include propositions of pure mathematics. It is entirely customary for alternative empirical hypotheses to draw on the same body of pure mathematics.[25] In such cases, the observations are used to test (H_1 & A) against (H_2 & A), but are not used to test the auxiliary assumption A against an alternative, since none was formulated in the first place. Cases of this sort involve epistemological asymmetry "by default," so to speak; they therefore differ from the case of Newman and his tuberculosis test in which alternative hypotheses and alternative auxiliary assumptions are both on the table for consideration.

---

[19] See Richard W. Miller, *Fact and Method* (Princeton: University Press, 1999), and my "Bayesianism. Its Scope and Limits," in Richard Swinburne, ed., *Bayes's Theorem* (New York: Oxford, 2002), pp. 21–38.

[20] I have emphasized how the likelihood approach allows one to avoid considering prior probabilities. However, there is a quantity that the likelihood approach requires one to consider, that Bayesian treatments (of the kind described in footnote 7) do not. When all four conjunctions of the form ($\pm H$ & $\pm A$) are simple statistical hypotheses, the likelihood approach will consider the quantity $\Pr(O \mid \text{not}H \,\&\, \text{not}A)$.

[21] See Isaac Levi and Sidney Morgenbesser, "Belief and Disposition," *American Philosophical Quarterly*, i (1964): 221–32, and Levi, *Gambling with Truth* (New York: Knopf, 1967).

Even so, the likelihood concept throws light on cases in which scientists decline to construct a new theory (or decline to construct a new set of auxiliary assumptions). They often do so because they expect that the new construction would have low likelihood, relative to *all* the evidence. Newtonian theory exhibited excellent fit to lots of other data; it would have been a very tall order to construct an alternative theory that does a better job of handling the data on Uranus while still having high likelihood relative to these other observations. The assumption that there are just seven planets was much less enmeshed with other data sets, so it made sense for Adams and Le Verrier to have held on to Newtonian theory while attempting to revise the auxiliary assumption about the number of planets.[26] As is well known, the same strategy met with failure when applied to the problem of explaining Mercury's orbit. Einstein's general theory of relativity was able to solve the problem precisely because it fit the data that Newton's theory also fit, while fitting the data on Mercury better. The old auxiliary assumption, that there is no planet between Mercury and the Sun, turned out to be right all along, notwithstanding the fact that Le Verrier and others considered the possibility that an as-yet unobserved planet (Vulcan) is perturbing Mercury's orbit.[27]

---

[22] Although Newtonian theory does not use the concept of probability, it still is a mistake to think of the theory plus auxiliary assumptions as *deductively entailing* a prediction about what one should observe concerning Uranus's orbit. The reason is that the observation procedures used are subject to error, and these error characteristics need to be modeled probabilistically, just as was true for Newman's tuberculosis test.

[23] W.M. Smart, "John Crouch Adams and the Discovery of Neptune," *Occasional Notes of the Royal Astronomical Society*, XI (1947): 33–88.

[24] Adams and Le Verrier were able to *accommodate* the observed orbit of Uranus within the Newtonian framework by postulating an eighth planet, but did Uranus's orbit *confirm* the hypothesis that there is an eighth planet? It is tempting to answer this question in the negative and to insist that it was only the observation of *Neptune* that provided evidence. The broad epistemological question at issue here is whether using an observation to construct a hypothesis means that the observation fails to provide evidence for the hypothesis. See Christopher Hitchcock and Elliott Sober, "Prediction, Accommodation, and the Problem of Overfitting," *British Journal for the Philosophy of Science*, LV (2004): 1–34, for discussion.

[25] See my "Indispensability and Mathematics," *Philosophical Review*, CII (1993): 35–57.

Must a solution to the Duhem-Quine problem give scientists advice on whether they should formulate an alternative to the hypothesis H or an alternative to the auxiliary assumptions A when the conjunction (H & A) generates a failed prediction? I do not think so. Epistemology does not have the burden of predicting that Uranus's orbit should be handled in one way while Mercury's should be handled in another. It took an Einstein (namely, *the* Einstein) to discover this; there was nothing in the anomalous data and their relation to Newtonian theory that indicated what the facts would turn out to be. It is perhaps more reasonable for philosophy in this instance to remain on one side of the divide between the *context of discovery* and the *context of justification*.[28] The likelihood analysis describes how alternatives should be compared once they are formulated, not whether they are worth constructing in the first place.[29]

---

[26] This point concerning "enmeshment with other data sets" is an appeal to observational evidence, not to the extra-evidential considerations that holists think are essential.

[27] Whereas the Adams/Le Verrier approach to the orbit of Uranus involved "asymmetry by default," this was not the case with respect to later discussion of the anomalous perihelion of Mercury, in that modifications of the auxiliary assumptions and of Newtonian theory were both developed. See N. Roseveare, *Mercury's Perihelion from Le Verrier to Einstein* (New York: Oxford, 1982), and John Earman and Michel Janssen, "Einstein's Explanation of the Motion of Mercury's Perihelion," in Earman, Janssen, and John D. Norton, eds., *The Attraction of Gravitation: New Studies in the History of General Relativity* (Boston. Birkhäuser, 1993), pp 129–72.

[28] See Hans Reichenbach, *Experience and Prediction* (Chicago: University Press, 1938)

[29] The distinction between *context of discovery* and *context of justification* is distinct

In discussing Newman's test result, the hypotheses considered were exhaustive (assuming that Newman exists), but the range of alternative auxiliary assumptions was narrowly circumscribed. In both versions of the problem, I assumed that applications of a tuberculosis test are independent and identically distributed; the same pair of error probabilities applies each time someone takes a test and the error probabilities that apply when one person takes the test are independent of what the outcomes happened to be when others did so. But surely this too is a background assumption that is up for grabs; it is not immune from revision in the context of the Duhem-Quine problem. The different auxiliary assumptions we considered disagree about the *values* of two parameters, but they agree on how the problem should be *parameterized*. It is perfectly legitimate to consider alternative *models* of how the test procedure works, where different models parameterize the problem in different ways. This takes us to our next topic.

IV. MODEL SELECTION

Although the likelihood approach is enough to show that epistemological holism is false, I do not claim that it is able to handle all instances of the Duhem-Quine problem. The main limitation concerns the treatment of composite (nonsimple) statistical hypotheses whose likelihoods cannot be interpreted objectively.[30] For example, consider the hypothesis that two physical quantities—say, the temperature and pressure in a closed chamber of gas—are related linearly. Although a specific straight-line hypothesis (for example, $P = 4 + 3T + U$, where U is an error distribution) confers a probability on a given value for pressure, given a value for temperature, it is more puzzling how one should think about the likelihood of the weaker claim that the relationship is linear (that is, that there exist values of a and b such that $P = a + bT + U$). This is because the likelihood of the hypothesis of linearity (LIN) is a weighted average of the likelihoods of all possible straight lines ($L_1, L_2, \ldots$):[31]

$$\Pr(\text{Data} \mid \text{LIN}) = \sum_i \Pr(\text{Data} \mid L_i) \Pr(L_i \mid \text{LIN}).$$

If one has no objective basis for saying how probable this or that straight line is, conditional on (LIN), one will not be able to treat the likelihood of (LIN) as an objective quantity.

---

from the distinction between *rules for accepting and rejecting hypotheses* and *rules for saying which hypotheses are better supported by the evidence*. I draw the latter distinction *within* the context of justification.

[30] See M. Forster, "Bayes and Bust: Simplicity as a Problem for a Probabilist's Approach to Confirmation," *British Journal for the Philosophy of Science*, XLVI (1995). 399–429, and my "Bayesianism: Its Scope and Limits" (*op. cit.*).

[31] As an expository convenience, I represent the average likelihood of (LIN) as a discrete summation, not as a continuous integration

The term "model" is used in the statistics literature on model selection to refer to hypotheses that contain at least one adjustable parameter. The hypothesis of linearity is a model in this sense, but specific straight-line hypotheses are not. Perhaps the most widely used model selection criterion is the Akaike information criterion (AIC), proposed by Akaike (*op. cit.*).[32] AIC is nonBayesian; the goal is not to compute the probability of a model or its likelihood. Rather, AIC aims to provide an estimate of the model's *predictive accuracy*.[33] But how can (LIN), as opposed to a specific straight-line hypothesis, provide a prediction (either accurate or inaccurate) about the gas's temperature when the gas is raised to a particular temperature? The answer is that the model must first be fitted to a set of old data; the parameters a and b are estimated from that data, using the method of maximum likelihood estimation. By substituting these estimates for the adjustable parameters, (LIN) is replaced with the specific straight-line hypothesis L(LIN); this is the specific straight line that renders the old data maximally probable. One then draws a new data set from the chamber of gas and determines how well L(LIN) predicts this new data. The average performance of (LIN) in this prediction task—first being fitted to old data, then seeing how well the fitted model predicts new data—defines the model's predictive accuracy. Estimating the predictive accuracy of models is the goal; the next question is how we should go about attaining that goal. If we have just one data set at hand, how are we to use this evidence to judge how predictively accurate a model is?

An important lesson that scientists absorb from working with models is that making a model too complex will reduce its predictive accuracy. It is easy to get a model to fit the available data by making it sufficiently complex, but the price is often that the fitted model does a poor job predicting new data. This is not a brute fact in the life experience of scientists; rather, Akaike provided a mathematical framework that helps explain why overfitting tends to reduce predictive accuracy. Using some very general assumptions, Akaike proved a result concerning how the predictive accuracy of a model can be estimated. He showed that an unbiased estimate of a model's predictive accuracy can be obtained by looking at two of its properties—how well it fits the data at hand, and how complex it is (where complexity is measured by the number of adjustable parameters the model contains). Akaike's theorem is the basis for AIC, which assigns a score to a model that reflects both its fit-to-data and its simplicity.[34] By comparing the AIC scores of different models, one can estimate which models will make more accurate predictions.

---

[32] See also Y. Sakamoto, M. Ishiguro, and G. Kitagawa, *Akaike Information Criterion Statistics* (New York: Springer, 1986); and Kenneth P. Burnham and David R. Anderson, *Model Selection and Inference: A Practical Information-Theoretic Approach* (New York: Springer, 1998).

[33] See Forster and Sober, "How to Tell When Simpler, More Unified, or Less *Ad Hoc* Theories Will Provide More Accurate Predictions," *British Journal for the Philosophy of Science*, XLV (1994): 1–36

As an example of a model selection problem in which models can be viewed as conjunctions (and so the Duhem-Quine problem can arise), let us consider the task of phylogenetic inference.[35] The goal is to evaluate the plausibility of different phylogenetic trees. Are human beings more closely related to chimps than they are to gorillas, or is the tree topology something different? There are three bifurcating trees that need to be considered—$(HC)\,G$, $H(CG)$, and $(HG)\,C$—but none of these confers a probability on the data until some model of the evolutionary process is provided. There are several process models to consider. In the context of molecular evolution, the simplest model is that of T. Jukes and C. Cantor,[36] which assumes that each of the four nucleotides has the same probability per unit time of changing into any of the others. More complex models, such as the one due to M. Kimura,[37] allow different changes to have different probabilities.[38] When a tree topology is conjoined with a process model, its adjustable parameters may be estimated from the data, and the AIC score of the conjunction may then be computed, as depicted in Table 7. Nothing prevents these cell entries from exhibiting the same asymmetries we saw in the likelihood analysis of Newman's tuberculosis test. The data might favor one tree topology over the other, regardless of which process model is used, but fail to provide any robust indication of which process model is better. It also is possible for the data to provide information about both tree topology and process model while providing more information about one than it does about the other. There is nothing special about AIC in this regard; all model selection criteria can generate both qualitative and quantitative asymmetries. This shows that the solution to the Duhem-Quine problem that I am proposing does not require a commitment to likelihood as the one true way to interpret evidence.

---

[34] Does AIC play into the hands of holism by invoking simplicity? I would say not, in that the justification for AIC depends on empirical assumptions (though ones of great generality). In addition, I will be using AIC only as an example of a model selection criterion; *cross-validation* is another such criterion, and it involves no appeal to simplicity. As it happens, take-one-out cross-validation is asymptotically equivalent with AIC. See M. Stone's "Cross-validatory Choice and Assessment of Statistical Predictions (with Discussion)," *Journal of the Royal Statistical Society*, B xxxvi (1974): 111–47, and his "An Asymptotic Equivalence of Choice of Model by Cross-validation and Akaike's Criterion," *Journal of the Royal Statistical Society*, B xxxix (1977): 44–47.

[35] See my "The Contest between Likelihood and Parsimony," *Systematic Biology* (2004, forthcoming).

[36] "Evolution of Protein Molecules," in H. Munro, ed., *Mammalian Protein Metabolism* (New York: Academic, 1969), pp. 21–132.

[37] "A Simple Method for Estimating Evolutionary Rates of Base Substitutions through Comparative Studies of Nucleotide Sequences," *Journal of Molecular Evolution*, xvi (1980): 111–20.

[38] For a survey of the different models now on offer, see Roderic Page and Edward Holmes, *Molecular Evolution: Phylogenetic Approach* (Malden, MA: Blackwell, 1998), pp. 148–62.

TABLE 7

		Process Models	
		Jukes-Cantor	Kimura
Tree topologies	$(HC)G$	AIC_1	AIC_2
	$H(CG)$	AIC_3	AIC_4

V. CONCLUDING COMMENTS

How is the likelihood analysis described here related to a deductivist formulation of the Duhem-Quine problem in which we recognize that there are three choices (and not just two)—we can reject H_1, reject A_1, or both? If the conjunction (H_1 & A_1) entails O, then $\Pr(\text{not}O \mid H_1 \;\&\; A_1) = 0$. The four likelihoods we need to consider, relative to the observation (not$O$), are shown in Table 8. Of course, the Law of Likelihood does not make recommendations about acceptance and rejection, but merely describes the differential support that the evidence provides. However, if acceptance and rejection require an evaluation of evidence, then the relationship of these four likelihoods is relevant to deciding what to accept and what to reject. If H_2 *dominates* H_1 (that is, if $p_1 < p_3$ and $p_2 < p_4$), then the evidence favors H_2 over H_1 regardless of which auxiliary assumption is true. If, in addition, A_2 dominates A_1 (that is, if $p_1 < p_2$ and $p_3 < p_4$), then (H_2 & A_2) is the conjunction with the highest likelihood, and so likelihood considerations point away from both H_1 and A_1. In this circumstance, the likelihood inequalities provide no qualitative asymmetry between the observation's impact on the hypotheses and its impact on the auxiliary assumptions, though a quantitative asymmetry may nonetheless obtain. If dominance holds in one direction but not the other, there is a qualitative asymmetry.

TABLE 8

		Possible Auxiliary Assumptions	
		A_1	A_2
Hypotheses	H_1	$p_1 = 0$	p_2
	H_2	p_3	p_4

In both the likelihood approach and the model selection approach, all four conjunctions of the form $(H_i \ \& \ A_j)$ $(i,j = 1,2)$ must be considered.[39] This brings out a further difference between these analyses and Quine's holistic epistemology. As noted earlier, the holist recognizes that people manage to decide which conjunct to blame when a conjunction generates a false prediction, but contends that this decision must rely on extra-evidential considerations, such as simplicity or conservatism. The problem is typically formulated in terms of what a person *believes*—if you believe the conjunction $(H_1 \ \& \ A_1)$ and then find that conjunction refuted by the evidence, which conjunct should you abandon? Quine recommends a policy of "minimum mutilation"—you should impose the smallest change in your web of belief that suffices to restore consistency with the observations, where changing a more "central" belief is regarded as a larger modification than changing a belief that is more "peripheral."[40] It follows that if you believe $(H_1 \ \& \ A_1)$, there can be no reason for you to abandon *both* conjuncts and embrace $(H_2 \ \& \ A_2)$, if a more conservative reformation to either $(H_1 \ \& \ A_2)$ or to $(H_2 \ \& \ A_1)$ would manage to restore consistency. It is interesting that the Law of Likelihood and model selection criteria place no premium on minimizing change in which propositions you believe. What you happen to believe plays no role at all; in fact, *you* do not enter into the analysis. The question is entirely about the relationship of propositions to data and has nothing to do with people and their affections or mobility. If $(H_2 \ \& \ A_2)$ has the highest likelihood or the best AIC score, so be it.[41]

The falsity of epistemological holism does not mean that it is *never* true; the point is that it is *often* untrue. Indeed, it is possible to describe a circumstance that can arise within a likelihood framework in which holistic intuitions are vindicated. Suppose that dominance fails, both with respect to the comparison of hypotheses and also with respect to the comparison of auxiliary assumptions. That is, let $p_1 < p_2$, $p_3 > p_4$,

---

[39] I hope it is clear that my focus on *four* conjunctions is an expository convenience. Both the Law of Likelihood and model selection criteria can address any number of conjunctions, and the asymmetries I have described can arise in that larger comparative context as well.

[40] *Philosophy of Logic*, p. 7; see also "Two Dogmas of Empiricism," p. 44.

[41] Since Quine's principle concerns change in what one believes, whereas the Law of Likelihood and model selection criteria do not provide rules of acceptance, there is no formal incompatibility here. However, when likelihood is placed in the context of a full Bayesian framework with prior probabilities, it can turn out that $(H_1 \ \& \ A_1)$ is the conjunction with the highest prior probability whereas $(H_2 \ \& \ A_2)$ is the one whose posterior probability is greatest. The imperative to find the conjunction that differs minimally from the one with the highest prior probability and that is logically consistent with the new evidence is not a principle of Bayesian epistemology.

$p_1 < p_3$, and $p_2 > p_4$ in Table 8. The cell entries in the table will then have two peaks (at H_1 & A_2 and at H_2 & A_1) and two valleys (at H_1 & A_1 and at H_2 & A_2). If the two peaks have equal likelihood, the observations will not discriminate between them, even though each peak is better supported than the conjunctions that occupy the two valleys. The history of epistemology is peppered with examples of this type. I will mention two; the first is Cartesian, while the second is due to Hans Reichenbach.[42] In both cases, (H_1 & A_1) predicts O, but not$O$ is what you observe:

not$O$: There seems to be a printed page in front of me.
H_1: There is a salami (and not a printed page) in front of me.
H_2: There is a printed page (and not a salami) in front of me.
A_1: My senses are functioning normally.
A_2: An evil demon is causing printed pages to look like salamis, and vice versa.

not$O$: My measurement device indicates that the triangle I have just measured has an angle sum that exceeds 180°.
H_1: Space has zero curvature.
H_2: Space has constant positive curvature.
A_1: There are no universal forces.
A_2: Universal forces are in operation.

Given the observational outcome not$O$, likelihood allows you to discriminate between conjunctions in the same row and conjunctions in the same column, but not between the anti-diagonal entries. The same situation can arise in the context of model selection, if the AIC scores in Table 7 have the pattern of peaks and valleys just described in connection with Table 8.

The likelihood approach views theory evaluation as inherently comparative. The Law of Likelihood describes what an observation says about *two* hypotheses, but is silent on the question of what, if anything, it says about a single hypothesis taken on its own. Likelihoodism elevates that omission to the level of principle; when a hypothesis is tested, it must be tested *against* alternatives.[43] In addition, there is no single alternative that counts as the uniquely correct alternative to consider. There are many specific alternatives, and the Law of Likelihood may be asked for its assessment of all these competitions. This marks an important difference between likelihoodism and Bayesianism. Bayesianism also has its comparative element, but there is just one alternative to a proposition that Bayesians need to consider; this is simply the proposition's negation. The comparative character of likelihood assessments (and of model selection criteria as well) is central to the analysis proposed here of the Duhem-Quine problem: *whether holism is right in what it says about the bearing of evidence on a conjunction crucially depends on what the alternative conjunctions are.* In many scientific contexts, the competing hypotheses and the competing auxiliary assumptions are such that the evidence has an impact on one conjunct that differs from the impact it has on the other. However, there are other discrimination problems in which the evidence cannot penetrate from an assessment of conjunctions to an assessment of conjuncts. Holism has its place, but in the context of scientific inquiry it is, by far, the exception and not the rule.

---

[42] *The Philosophy of Space and Time* (New York: Dover, 1958)

[43] An obvious exception arises when a proposition entails an observational prediction that fails to come true. In this instance, there is no need to consider an alternative; *modus tollens* permits one to reject the proposition without further ado. However, in the very common case in which H, or the conjunction (H & A), confers a probability on O that is less than unity, the falsity of O cannot be interpreted until alternatives are considered. Likelihoodism therefore rejects "probabilistic *modus tollens*" (a.k.a. Fisherian significance testing), according to which a hypothesis should be rejected when an event occurs that the hypothesis says was very improbable. For criticisms of probabilistic *modus tollens*, see Hacking, Edwards, and Royall.

Quine famously opined that "any statement can be held true come what may, if we make drastic enough adjustments elsewhere in the system."[44] Quine's point was not just that it is *logically possible* to continue to believe a proposition in the face of apparently recalcitrant experience, but that *the evidence* does not say that this is a mistake. Decisions about retention or rejection are dictated by nonevidential considerations. I suspect that the allure of this form of holism derives from a hypothetico-deductive view of theory testing. If an observation O confirms a proposition P precisely when P entails O and O turns out to be true, and O disconfirms P precisely when P entails O and O turns out to be false, then epistemological holism is correct when the conjunction (H & A) entails O, but neither conjunct does. In this circumstance the observational outcome confirms or disconfirms the conjunction, but not the constituent conjuncts. However, it is abundantly clear that confirmation and disconfirmation can be mediated by *non*deductive relationships. In the examples discussed here concerning tuberculosis diagnosis and phylogenetic inference, the conjunctions considered were each consistent with the observational outcome without entailing it, but that does not mean that the evidence failed to discriminate among them. Evidence can do more than hypothetico-deductivism imagines. It is that extra power that undermines epistemological holism.

ELLIOTT SOBER

Stanford University
University of Wisconsin/Madison

---

[44] "Two Dogmas of Empiricism," p. 43.

COMMENTS AND CRITICISM

EVIDENCE, EXPLANATION, AND EXPERIENCE: ON THE HARDER PROBLEM OF CONSCIOUSNESS[*]

A new type of problem is taking hold in the philosophical debate about consciousness, one that adds to the so-called "hard problem of consciousness."[1] It concerns our epistemic relation to the possible phenomenal states of creatures that are different from us in their physical realization. Ned Block has recently given a thorough account of this "harder problem of consciousness,"[2] and David Papineau, David Chalmers, Joseph Levine, and Thomas Nagel have discussed similar but less substantial versions of this type of problem.[3] It arises for naturalistic phenomenal realists, that is, those who have good but defeasible reasons for believing that consciousness has a scientific nature, and who on the basis of first-person grasp of consciousness believe that phenomenal experiences are instantiations of real properties, not something that can be conceptually reduced away in the manner of analytical functionalism.

Block's harder problem of consciousness (HPC, hereafter) is that naturalistic phenomenal realists face an epistemic tension: they have no conception of a rational ground for believing that other creatures, who do not relevantly share our physical nature, are conscious or not. Even if they assume that the other creature is conscious, they have no conception of a rational ground for believing that consciousness in this creature has a physical basis, even though physicalism is the default view. Furthermore, it is the subjective default that other crea-

---

[*] Thanks to the audience at the Annual Meeting of the Danish Philosophical Association, February 2003, especially to Ned Block; and to the audience at a seminar at the Philosophy Program, Research School of Social Sciences, Australian National University. The research for this article was supported by a grant from the Danish Research Council for the Humanities to the research project *Naturalised Mind—Cognisant Nature* (www.namicona.au.dk).

[1] See David Chalmers, *The Conscious Mind* (New York: Oxford, 1996).

[2] "The Harder Problem of Consciousness," this JOURNAL, XCIX, 8 (August 2002): 391–425. Page references are to this article unless otherwise noted.

[3] Papineau, *Thinking about Consciousness* (New York: Oxford, 2002), chapter 7; Chalmers, "Availability: The Cognitive Basis of Experience?" in Block, Owen Flanagan, and Güven Güzeldere, eds., *The Nature of Consciousness* (Cambridge: MIT, 1995), pp. 421–24; Levine, *Purple Haze: The Puzzle of Consciousness* (New York: Oxford, 2001), chapter 3; and Nagel, "The Psychophysical Nexus," in Paul Boghossian and Christopher Peacocke, eds., *New Essays on the A Priori* (New York: Oxford, 2000), pp. 433–71, footnote 21.

0022-362X/04/0105/242–54

tures are not conscious, but the phenomenal realist must leave it an open question whether they are.

Block discusses HPC in terms of a conditional whose antecedent is the conjunction of naturalism, phenomenal realism, and anti-skepticism about other minds (the latter is necessary to distinguish HPC from the other minds problem). The consequent is the epistemic tension. I discuss two aspects of Block's arguments that this conditional is true.

First, I assess the argument to the conclusion that naturalistic phenomenal realists have no conception of a rational ground for belief in disjunctive physicalism. I argue that it is possible to resist the conclusion because, given a plausible notion of inference to the best explanation, the argument begs the question (section I).

Second, I assess the argument that naturalistic phenomenal realism produces an epistemic obstacle to discovering phenomenality in other creatures. This argument ignores the priority of explanation over discovery in inferences to the best explanation (section II). I further explore the relationship between evidence, explanation, and conscious experience: some of the hardness of HPC may stem from doubts about the occurrence of the evidence, or from doubts about the interpretation of the evidence. But the first of these doubts is misplaced, and the second may be overcome, again with help from the notion of inference to the best explanation (section III).

My criticism is not directed at the grounding intuition that there are epistemic obstacles in the study of consciousness; it is pretty clear that we have this intuition. It is directed at the arguments that Block mounts, on the basis of this intuition, to describe the nature and consequences of our current epistemological predicament.

I. THE ARGUMENT FOR META-INACCESSIBILITY OF DISJUNCTIVE PHYSICALISM

Throughout, Block makes use of the science fiction case of Commander Data, who is a merely superficial functional equivalent to us, equivalent to us in respect of folk psychology and what folk psychology entails, but not in other psychological and neuropsychological functions. Importantly, however, Data is unlike us in the physical nature and organization of the control mechanisms of the folk-psychological functions (401–02).

We have good but defeasible grounds for believing that Data is conscious because he acts like us, and we act the way we do in part because we are conscious. The grounds are defeasible because we might find that Data's physical constitution shares none of our neural correlates of consciousness (402–03).

One of Block's central premises is that we have no conception of a rational ground for the belief that Data is or is not conscious. He is stipulated to be physically different from us, so any science of consciousness that is based on us will be powerless to determine whether he is conscious or not. And Block argues further that we cannot base a science of consciousness on him unless we already know he is conscious (I return to this argument in section II). So we have no idea what would be evidence for the belief that Data is conscious or not; Data's consciousness is meta-inaccessible (405).

If the phenomenal realist is a physicalist who thinks that consciousness has a deep, nondisjunctive physical basis, then Data is not conscious because he does not have this basis. But this seems unduly chauvinist. A weaker disjunctivist physicalism allows that Data can be conscious. Disjunctivism says that consciousness is the heterogeneous physical property of having either our or Data's realization of our shared functional state.

Disjunctivism is attractive to the naturalistic phenomenal realist because it allows that we and Data can have overlapping phenomenality with a common, albeit disjunctive, scientific basis. The problem is that, even though it is a default that consciousness has a scientific nature, we have no conception of a rational ground for inferring that disjunctivism is true. It is very odd that someone who believes that consciousness is real and has a scientific nature must also acknowledge that there is no conception of what the evidence for the naturalist position would be (407–08, 413–14).[4]

Block's argument to this conclusion begins by noting that one needs to arrive at disjunctivist physicalism via inference to the best explanation. The identification of phenomenality with the one disjunctive property of our or Data's realizations must contribute to the best explanation of facts about phenomenality, such as why there is phenomenal overlap between us and Data (410–12). This is similar to the way in which the identification of water with H_2O contributes to the best explanation of facts about water, such as that it expands when freezing (409–10). Of course, this only makes sense if it is assumed that Data is conscious: there is no best explanation of the phenomenal overlap, if he is not conscious (411).

There are now two arguments against arguing via inference to the best explanation to the conclusion that consciousness is a heterogeneous physical disjunction.

---

[4] Block also considers another physicalist doctrine, superficialism (412), which I will ignore here.

The argument about fundamental difference. The first argument in effect begins by stipulating that our and Data's physical realizations are fundamentally different in a sense that makes us balk at a disjunctive explanation of facts about phenomenality. When it comes to two fundamentally different properties, distinct explanations are better than one disjunctive explanation. Block illustrates the point by saying that we could, in theory, explain solid-like formation in terms of crystallization or continuous hardening into an amorphous substance (which would make glass a solid). But we reject the disjunctive explanation and prefer two distinct explanations because the properties of crystallization and continuous hardening into an amorphous substance are fundamentally different. Block does not specify the basis for judgments of fundamental difference, he just notes that we have a vague notion of it and then proceeds to the above stipulation (411).

Block's argument leads up to the conclusion that disjunctivism is meta-inaccessible: that we have no conception of a ground for rational belief in disjunctivism. If our and Data's realizations are fundamentally different, then there can be no argument via inference to the best explanation to a physical disjunction (though disjunctivism may of course still be true).

Nevertheless, this use of the notion of fundamental property difference is questionable. To begin with, it seems doubtful whether a merely intuitive notion of fundamental difference should guide our grounds for rational belief. Intuitively (that is, in the absence of knowledge of the scientific truth), diamonds and graphite are fundamentally different, but they are surely fundamentally similar, since science tells us they are different crystallizations of carbon.

Then there is the case of solid-like formation. It is my understanding that water can freeze into a glassy, amorphous solid when it is cooled rapidly,[5] just as in the case of window glass, which is sand that is melted and rapidly cooled. So here we have solid-like formation—freezing into glassy ice—that is explained in terms of continuous hardening, not crystallization. This opens the possibility of a disjunctive explanation, at least by the indicators employed by Block: glassy ice is *not* a supercooled liquid (glassy solids are more viscous, though still metastable, counterparts of supercooled liquids[6]), and its formation *is* a kind of freezing. But then a reasonable case can be made that we have an explanation of solid-like formation in terms of a disjunction of intuitively fundamentally different properties, namely, continuous hardening-or-crystallization. So intuitive fundamental difference is not in itself sufficient to close off the feasibility of disjunctive explanation.[7]

---

[5] Philip Ball, *H_2O: A Biography of Water* (London: Weidenfeld and Nicholson, 1999), p. 187. Block notes this phenomenon as well, and says we do not regard it as freezing (411).

[6] Osamu Mishima and H. Eugene Stanley, "The Relationship between Liquid, Supercooled and Glassy Water," *Nature*, CCCXCVI (1998): 329–35.

The point is not that the discovery of glassy ice straight off should make us accept the disjunctive explanation; it is that the discovery makes it reasonable to suggest that the explanation is disjunctive, in spite of the intuitive fundamental difference. An epistemically defensible notion of fundamental property difference may therefore depend on empirical work and on explanation, and is thus not so easily up for uncontroversial stipulation. In particular, it seems plausible that judgments of fundamental property difference and similarity are, at least in part, based on how well assumptions of fundamental similarity or difference contribute to overall explanatory integration.

The underlying thought is that we would not rationally care much about fundamental similarity and difference unless it had an impact on the overall explanatory picture. If we achieve the best explanatory integration of a wide range of phenomena by one grouping of properties rather than another, then we go with the first. Allowing the disjunctive explanation of solid-like formation may be messier, and less conducive to overall integration than going with the two distinct explanations (it may be less messy in the overall scientific picture to count glass as a liquid). Conversely, once we find out about the glassy solid state of water we might well decide that it is the disjunctive picture that is less messy. It is hard to think of an alternative basis for judgments of fundamental property difference, short of some kind of aesthetic preference that we can hardly allow to inform rationally these epistemic issues.

On this view, the stipulation that our and Data's realizations are fundamentally different is not independent of issues concerning best explanation. Rational judgments of fundamental property difference come *after* we evaluate which explanations are best; we cannot avail ourselves of them first and then use them to tell us which explanations are best. But then the best way to make sense of the stipulation presupposes that there is no argument from inference to best explanation to a physical disjunction. This begs the question, for that is the conclusion at which Block is aiming.

---

[7] We could take this to show that *solid-like formation* is a merely nominal category, like *jade*. But this would beg the question since it in effect says that only nominal categories can have disjunctive explanations

Objection: it is possible to obtain scientifically informed views on fundamental property difference on the basis of all *other* explanations than explanations of consciousness, and then to pass judgment on the fundamental difference of our and Data's realizations. *Reply:* this is unduly conservative: even if they were judged different on the basis of other explanations, the best-making qualities of explanations of consciousness may override this judgment (just as the best-making qualities of the explanation of solid-like formation may override previous judgments, once we learn about glassy solids).

Objection: Block's argument explicitly concerns our epistemic predicament in a pre-explanatory situation (406, 424); so it begs the question against Block to invoke what we may say about property difference once explanations become available. *Reply:* the argument is not that HPC will disappear once explanations of consciousness become available, it is rather that Block faces a dilemma: be irrationally guided by pre-explanatory judgments about fundamental property difference or beg the question by presupposing that there is no best disjunctive explanation.

The argument about bad and good questions. Block's second argument against arguing via inference to the best explanation to the conclusion that consciousness is a heterogeneous physical disjunction concerns the neatness of best explanations. Considerations of neatness say that best explanations should solve more explanatory puzzles, and not give rise to bad explanatory puzzles. The theoretical identification of water and H_2O, and heat in gases and mean molecular kinetic energy, prevent bad questions about why it is that water is *correlated* with H_2O and why it is that heat is *correlated* with mean molecular energy from arising; there is no explanation of why things correlate, if they are identical. According to Block, this is part of the reason why we accept these identities.

Now consider the question of why we overlap with Data in this or that phenomenal respect. This question does not arise if we accept disjunctivism. If there is phenomenal overlap, then this just *is* our having the disjunctive property of our or Data's realization—and identities do not have explanations. Block now claims that the question does not seem bad, and therefore should not be ruled out, so, accordingly, we should not accept the theoretical identification. (This is not evidence that Data is not conscious since, for the phenomenal realist, it is still an open question whether he in fact is.) If inference to the best explanation cannot do the job, then there seems no source of a conception of a rational ground for belief in disjunctivism.

We have to ask, however, what makes us deem that certain why-questions are good and others bad. In an argument about the epistemic viability of physicalism, we cannot let ourselves be rationally guided merely by how questions *seem*; that is too arbitrary. Moreover, it seems plausible that such judgments can be wrong: what we once deemed a good question may later transpire as a bad question, and vice versa. My view is that it is explanation that can make us change our minds about the quality of why-questions. We say that the question about why it is that water is correlated with H_2O is bad because the best explanation of facts about water involves acceptance of the theoretical identification of water and H_2O. If we had no such explanation it is very plausible that it would be deemed a good question, at least by some (such as staunch phlogiston-believers who are just coming to terms with Lavoisier's notion of *oxygen*).

Some explanations rule out some questions as bad and other explanations rule out other questions. But it is only after we have the overall explanatory picture that we can make an informed judgment about which questions were good and which bad. What this shows is that the ruling out of bad questions, in Block's sense, is not an independent best-maker for explanations, and thus should not in itself be used to plead acceptance or rejection of theoretical identities. But then, under the assumption that there is as yet no explanation, arguments based on these notions are not available, unless the question is begged in favor of there not being a best disjunctive explanation.

Objection: What you argue is that as long as there is no explanation of consciousness, there will be no resolution of these problems, but is that not precisely the situation Block describes—namely, that in our current epistemic situation the questions about physicalism about consciousness are meta-inaccessible? *Reply:* Block is arguing *for* the claim that physicalism is meta-inaccessible; it is not a premise in his argument. My argument is that this argument does not go through without begging the question. The arguments in this section do not concern the *premise* that we have no conception of a rational ground for belief that Data is conscious or not.

II. EXPLANATION AND DISCOVERY

Block's premise about the meta-inaccessibility of Data's consciousness plays an important role in the overall argument. The argument is that any science of consciousness that can generalize to other creatures must be based in part on them in the first place, but that this cannot be done unless we already have discovered whether they are conscious or not (406–07). The premise about Data prevents this discovery. So, as things stand, there will be no generalizable science of consciousness.

Behind this argument is the following epistemic principle about the priority of discovery over explanation:

(DE) There can be no explanation of a phenomenon unless it is already discovered that it occurs.

I think DE gets some of its intuitive plausibility from the following metaphysical principle:

(OE) There can be no explanation of a phenomenon that does not occur.

For example, there can be no explanation of the occurrence of global peace if it does not occur (though of course there can be explanations of the nonoccurrence of a phenomenon, like the nonoccurrence of global peace). OE is weaker and less controversial than DE because it does not require any discovery of the occurrence of the phenomenon; it is consistent with a range of epistemic states vis-à-vis the phenomenon, and with discovering its occurrence after having explained it.

In his influential discussion of inference to the best explanation, Peter Lipton notes that the point of explaining is not just to explain *why* something is the case; in the shape of inference to the best explanation it is also our primary tool for discovering *what* is the case.[8] One of Lipton's examples illustrates this. Consider the red shift observed in the light spectrum of some stars. This was taken to be evidence of some unknown, undiscovered phenomenon. Various hypotheses can be proposed to explain the evidence, and the one that best explains the evidence is the one we come to believe, in this case that the star recedes with a certain velocity. So we used best explanation to discover something about the world. The principle is:

(ED) Sometimes best explanation can lead to discovery.

This principle does not contradict DE. DE says that the explanandum phenomena must already be discovered. In the recession example, the explanandum phenomenon is the red shift, which *is* discovered. The discovery that is relevant to ED concerns the explanans, that is, the recession. The question is, however, whether ED will help circumvent the role that DE plays in Block's argument. Can explanation of various kinds of phenomena lead to discovery of consciousness in other creatures?

What is the best explanation of the evidence of folk-psychological functionality in Data? We would normally reason from like effects to like causes and say the explanation is the occurrence of phenomenal states. But this inference is not sound here because we know that Data's realization is different from ours, and this defeats the argument by analogy.

---

[8] *Inference to the Best Explanation* (New York: Routledge, 1991), pp. 57–58, 68–69.

Consider instead the possibility of inferences to hypotheses that explain the organizing principles of the physical properties of Data's silicon brain and how they give rise to the folk-psychological functionality. One of these hypotheses may explain some of this evidence partly in terms that clearly identify the phenomenal property of pain with a particular pattern of silicon brain activity. Another may explain it partly in terms of the identity of a nonconscious, merely cognitive property and that particular pattern of silicon brain activity. We should accept the hypothesis that best explains the evidence. If this hypothesis is the one that identifies phenomenal properties with silicon brain properties, then that is the one we should believe. That is, best explanation can lead to the discovery of phenomenality.

Objection: this discussion of HPC presupposes that we have a conception of explanations of how a physical property can be identical to a phenomenal property but HPC is explicitly set in a pre-explanatory context (406, 424). *Reply*: this is irrelevant to the issue at hand, which concerns an argument about the relative priority of discovery and explanation. It begs the question in favor of the thesis that discovery must precede explanation if we cannot assume for the sake of argument that explanations are available.

Objection: HPC is not a principled problem, it does not attempt to say that we *could* not have a conception of a rational ground for belief in Data's being conscious or not. Rather, it is contingent on the fact that we have not yet managed to explain how a physical property can be identical to a phenomenal property, and thus it merely argues that our current epistemic predicament about consciousness is of a certain nature and has its origin in naturalistic phenomenal realism. It is no criticism of this view to note that HPC is resolved once we arrive at such explanations. *Reply*: it is one thing to note our strong intuition that, if confronted with Data, we would have no conception of a rational ground for belief that he is conscious or not. I do not deny that we have this intuition. It is another thing to *argue* that we need to discover phenomenality before we can have a generalizable science of consciousness, or indeed that disjunctivism is not epistemically viable. Perhaps our intuitions support these conclusions (though I am not sure about that), but it does not follow that the arguments leading up to the conclusions are convincing. My criticism so far is that these arguments are not convincing. So we are left with the intuitions.

III. DOUBTS ABOUT THE OCCURRENCE AND INTERPRETATION OF THE EVIDENCE

The argument in the preceding section establishes that explanation can lead to the discovery of consciousness in other creatures, and thus we can circumvent Block's claim that there cannot be a science of consciousness for creatures like Data. It is, however, possible to resist this conclusion by changing the context of doubt so that it no longer concerns what phenomena we have and have not discovered, but instead concerns our justification for believing in the occurrence of the evidence. This kind of doubt may be the source of the hardness of the harder problem, and it does indeed lead to a very hard problem, but the underlying doubt is misplaced. Another kind of doubt, about the interpretation of the evidence, is not misplaced but can be dealt with by once again employing properties of inference to the best explanation.

Doubts about the occurrence of the evidence. Here is how the doubt arises. We can grant that best explanation can lead to discovery, but only if there are no doubts about the occurrence of the evidence for the explanation. For example, explanation of a red shift can lead to discovery of recession, but only if there are no doubts about the occurrence of the red shift (was it a genuine red shift, or an artifact of our measuring devices?). Similarly, the doubt goes, "silicon brain science" explanations of the evidence concerning the organizing principles of the physical properties of Data's silicon brain and how they give rise to the folk-psychological functionality can lead to discovery of consciousness, but only if there are no doubts about the occurrence of the evidence. But there are such doubts, for perhaps the organizing principles are designed to mimic genuine folk-psychological functionality and are therefore no more evidence of phenomenality than a tape-recording of fabricated introspective reports.

We cannot respond by claiming that if there *is* a best explanation in terms of phenomenality then there are no grounds for doubting the evidence. This kind of argument leads to vicious circularity in cases where the phenomena that are being explained are themselves essential parts of the evidence for the explanation. If someone doubts the occurrence of the red shift in the light spectrum of a star, then we cannot justify belief in its occurrence with reference to the best explanation (that is, in terms of recession), since the red shift is evidence for the recession. Similarly, if someone doubts the occurrence of the evidence concerning the organizing principles of Data's brain and the functionality, then we cannot justify belief in its occurrence with reference to a best explanation (that is, in terms of phenomenality), since it is an essential part of the evidence for that explanation.

This is a general feature of scientific explanation, and in particular of inference to the best explanation. Some explanations are self-evidencing, that is, the phenomena they explain are themselves part of the evidence for the explanation. There is circularity here, but it is benign. It turns vicious, however, when the aim is to use best explanation to establish the occurrence of the evidence.[9]

Hence, in the context of doubts about the occurrence of the evidence, it seems there is no way of forming a conception of a ground for rational—that is, noncircular—belief in Data's being conscious or not, even if there are available explanations that identify phenomenal properties with some physical properties of his brain! In this version, the problem posed by Block's premise about the meta-inaccessibility of Data's phenomenality is therefore very hard. In contrast to the hard problem, it will not immediately go away once we have a conception of what explanations of phenomenality will be like.

Doubts about the interpretation of the evidence. Luckily, this is not a good way to characterize the doubt that one may have about the evidence. For in a very real sense there is no doubt about the occurrence of the evidence. Data's folk-psychological behavior, and the silicon circuitry in his brain, is in principle there for all to observe. We have no reason to introduce some fault in our methods for acquiring this kind of evidence.

I think there is a better way to capture the relevant doubt about the evidence, but that the problem posed by this doubt is solvable such that we can have a plausible conception of a rational ground for belief, also post explanation. It is possible to view the problem as concerning the interpretation of the evidence rather than the occurrence of the evidence. Sometimes people disagree about the acceptability of various hypotheses because they disagree about the standards of interpretation of the evidence, not because they disagree about whether the evidence confirms the hypothesis. Such disputes seem intractable because any further evidence that gets drawn in to settle the matter is itself subject to the different standards of explanation (which is why this kind of dispute also plays a role in Thomas Kuhn's notion of incommensurability).

By once again employing the resources of inference to the best explanation it is nevertheless possible to decide such disputes.[10] The sense than that familiar in deductive logic. "The probability that $h > r$" and "*the probability that* $h \leq r$" can be used to express attitudes that are inconsistent or incoherent in the extended sense favored by Frank Ramsey and Bruno De Finetti. To be committed to such judgments in the same context violates canons of rational probability judgment even though, as both of these authors rightly pointed out, such attitudes are neither true nor false. Perhaps, choosing x from $\{x,y\}$ and choosing y from $\{x,y,z\}$ can be inconsistent in a sense that implies irrationality after all.

---

[9] Carl G Hempel, *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science* (New York: Free Press, 1965), pp. 370–72; Lipton, pp. 26–27.

[10] Cf Lipton, pp. 71–72.

Sen's main point nonetheless stands. In order for a choice function C to be subject to rational appraisal, the judgments of admissibility represented by the choice function for the hypothetical situations calling for decision should be constrained by the agent's value commitments and beliefs. The constraints have often been required to satisfy the following conditions:

(i) There must be a "universal set" Ω of candidate options such that the choice function C is a function from nonempty, finite subsets of Ω to finite subsets of Ω such that $C(S) \subseteq S$.
(ii) Decision maker X's value commitments and beliefs determine a value structure $V(\Omega)$ for Ω representable by a weak ordering.[4] X's value commitments and beliefs also determine such a value structure $V(S)$ for every finite, nonempty menu $S \subseteq \Omega$.
(iii) The preference is ***menu independent*** with respect to Ω (see 172). Given any nonempty subset S of Ω, the value structure $V(S)$ of S when S is the restriction of $V(\Omega)$.[5]
(iv) $C(S)$ is the set of options in the option menu S that are optimal according to $V(S)$. $C(S)$ is thus a ***menu independent choice function*** (173).

Conditions (ii) and (iv) will be weakened later. For the present illustrative purposes, we shall remain with (i)–(iv) as stated. If conditions (i)–(iv) are met, the choice function must satisfy the requirements of property α mentioned above and the following property β: If $x,y \in S \subseteq T$, $x,y \in C(S)$, then $x \in C(T)$ if and only if $y \in C(T)$. Sen's critique of internal choice consistency emphasizes that whether or not conformity with α and β for specific Ω and C is rationally required or not will depend upon whether the decision maker X's value commitments and beliefs require satisfaction of (i)–(iv) for the given Ω and C. If they do, X's choices violating α are inconsistent *with those value commitments and beliefs* in a sense that is offensive to reason. However, there is no principle of rationality that entails that X endorses those value commitments and beliefs. And there is no principle of rationality that mandates *internal* choice consistency stipulating that X conform in the choices made to α regardless of X's value commitments and beliefs. As Sen says, without appeal to what the agent is "trying to do," no universal set is specified and no choice function is singled out with respect to which preference and choice function are menu independent.

---

[4] The weak preference relation xRy (x is as good as y) is reflexive (xRx), transitive ($xRy \& yRz \rightarrow xRz$) and complete ($xRy \vee yRx$). Strict preference xPy holds if and only if $xRy \& {\sim} yRx$. Equipreference xIy holds if and only if $xRy \& yRx$. That is to say, the ordering is not only *complete* but *connected*. $xPy \vee xIy \vee yPx$.

[5] Such preference is "robust" according to Levi, *Hard Choices* (New York: Cambridge, 1986), 6.3.

Sen is entirely right to object to rationality as internal choice consistency. Yet, there is one aspect of his discussion with which I have some difficulty. In marshalling his illuminating counterinstances to internal choice consistency, Sen neglects to indicate how a universal set is constrained by what the agent is "trying to do"—that is, by the agent's value commitments and beliefs. It seems to me that when the agent's value commitments and beliefs are taken into account, the issue does take on a nontrivial significance.

Sen has been interested in the question of "positional choice," in particular, for a long time due in part to its relevance for social choice theory where several voting procedures are positional. He examines several illustrations outside of social choice theory exemplifying the same process. Here is a useful example.

X is presented by his host with a basket of fruit consisting of mango m_1, apple a_1, and apple a_2. X desires to eat a mango but courtesy requires that he decline the only mango in the basket and so he chooses one of the apples.

Had the basket, counter to fact, contained another mango m_2, X would have chosen one of the mangos—perhaps, m_1. The set of potential options Ω includes taking mango m_1, mango m_2, apple a_1, and apple a_2 and refusing any fruit. The menu of options available to X is the set $S = \{m_1, a_1, a_2,$ refuse$\}$. X's "all things considered" evaluation ranks a_1 and a_2 as best, refusal next, and m_1 at bottom. $C(S) = \{a_1, a_2\}$. But in a case where there are several mangos and several apples, X may choose a mango. $C(T) = C(\{m_1, m_2, a_1, a_2,$ refuse$) = \{m_1, m_2\}$. Property α has been violated.

Sen points out that there is nothing irrational here. The choice function fails to be menu independent as (iv) requires and the preference fails to be menu independent as (iii) demands. Under the circumstances, rationality does not require conformity with property α. X's value commitments include the injunction never to take the only mango from a basket when others are present. And that injunction does not change when the options are all those in Ω. But the value structure for S is not embedded in the value structure for T. This menu dependence of the preference is explainable as due to a difference in X's information (and, hence his beliefs) about his options in the two scenarios. Ω has not changed. X's value commitments have not changed. X's information about X's options has changed. In the first scenario, choosing m_1 is known to be equivalent to choosing the only mango. In the second scenario, choosing m_1 is known to be a case of choosing one of several mangos. This change in information alters the change in the evaluation of the options. Hence, the choice function defined over finite nonempty subsets of Ω fails to satisfy α.

It is possible, however, to describe X's judgments of admissibility using a choice function C' of menus carved out of a different universal set Ω' that does, according to X's value commitments and beliefs, satisfy α. It can be argued that the options in the first case are {choosing the only mango, choosing one of several apples, refusal}. Perhaps, in deference to those who insist that there are four options and not just three, we may consider {choosing the only mango—to wit, m_1, choosing one of several apples—to wit, a_1, choosing one of several apples—to wit, a_2, refusal}. In the second case, the options are {choosing one of several mangos, choosing one of several apples, refusal} or the expansion to five options along the lines just indicated. Ω' is a superset of the union of these two sets. Notice that Ω' is different from Ω. X intends to do something different in choosing the only mango, namely m_1, than X intends in choosing one of several mangos, namely m_1.

It may perhaps be wondered how these two options can be considered part of the same universal set Ω'. To do so is to think it entertainable at least hypothetically that X can face a menu consisting of just this pair of options. This seems absurd. But it is surely coherent to consider a scenario where X is offered a choice between selecting a mango from a basket containing only one and nothing else and a basket containing two and nothing else. If one keeps in mind that the object of choice is an action of some type or kind and that the universal set is a set of such types of action, there should be nothing problematic here. With Ω' as the universal set and X's value commitments and information, the preference over Ω' is menu independent as is the choice function C' defined over subsets of Ω'. This choice function has property α.[6]

Sen thinks that such redescription strategies trivialize the question of choice consistency. But what gets trivialized? No matter what the universal set is over which the choice function is defined, compliance with property α may fail as a matter of fact. Nor is such conformity rationally required for choice functions defined over arbitrary universal sets. It is not required for choice function C defined over Ω. It ought to be required for C' defined over Ω'. That is because *X*'s value commitments and beliefs entail that conditions (i)–(iv) are satisfied. And when conditions (i)–(iv) are satisfied, α ought also to be satisfied. But it is far from trivial that *X* has the value commitments and beliefs that support conditions (i)–(iv) and a choice function satisfying α. What is trivialized?

---

[6] These remarks do not refute theorem 6.3 of Sen, on 163 There the universal set is held fixed. What Sen overlooks is the possibility of adjusting the universal set so as to obtain a menu independence over the preference set.

In the example under discussion, consideration of the value structure over Ω' shows that at least one universal set of potential options and a value structure consonant with *X*'s value commitments and beliefs over that universal set can be constructed such that both menu independence and choice function independence and, hence, choice consistency are required.

Both Ω and Ω' can be used to describe *X*'s choices. But in this example Ω' is better suited to represent *X*'s propensities to deliberate choice. Ω' utilizes descriptions of *X*'s choices relevant to the value commitments of *X*. Any further specification of the options Ω' would be irrelevant to determining whether *X*'s choices conform with *X*'s value commitments and beliefs. If the options are represented as they are in Ω, relevant information is left out. In order to obtain a redescription that produces the information relevant to the evaluation of the options, we need to derive the descriptions in Ω' appealing to *X*'s epistemic state. We can, indeed, do this. But there is nothing trivial about the fact that the specification of the options in Ω' provides the information relevant to promoting *X*'s goals and values whereas the specification of the options in Ω does not.

The moral of the story is this: whether so-called choice consistency conditions ought rationally to be satisfied or not depends upon the universal set relative to which choice functions are defined and the value structure defined over it. Roughly speaking, the specification of the space of potential options over which a choice function is to be defined ought to be one that includes for each option the information relevant to its appraisal according to *X*'s value commitments and beliefs.[7] Call this the condition of *Universal Set Relevancy*. In our toy example, Ω used by Sen to supply illustrations of failures of choice consistency fails Universal Set Relevancy. Ω′ on the other hand does meet the condition. *X*'s choice function conforms with choice consistency requirements relative to Ω′.[8]

---

[7] Sen offers his counterinstances to α in the context of a study of "choice behavior" rather than in the setting of "normative choice theory" (159 n.2). But a study of choice behavior that does not describe choices in a manner relevant to assessing the rationality of such behavior in terms of the goals and values of the decision maker ought, according to Sen, to be of doubtful usefulness.

As far as Sen's argument goes, therefore, rational agents might be expected to satisfy *relevant* choice consistency conditions even though they need not satisfy *internal* choice consistency conditions. Obeying choice consistency relative to universal sets that satisfy Universal Set Relevancy may prove disconcerting to those who engage in empirical investigations of consumer behavior by seeking to rely on choice behavior exclusively. Unless we are confident that changes in prices and income do not alter the "indifference maps" of individual consumers, a given space of commodity bundles may not qualify as a universal set of potential options.

There are situations where the relevant universal set is generated according to some definite procedure by a core set *S* of options that may or may not be the menu of available options confronting the decision maker. A familiar technique due to John von Neumann and Oskar Morgenstern identifies Ω with the *Mixture Set* $M(S)$ of *value neutral* lotteries using stochastic devices to select an element of *S* to implement. On the supposition that the preference over $M(S)$ in which the preference over *S* is embedded satisfies the von Neumann-Morgenstern axioms, numerical values may be assigned members of the mixture set $M(S)$ that are expectations of the values assigned members of *S*. It is important to recognize, as von Neumann and Morgenstern themselves recognized, that such conditions preclude assigning any value to the process of implementing a chance mechanism to select an element of *S*. If the von Neumann-Morgenstern method for eliciting cardinal utilities is to be deployed, a mixture of pure options is to be evaluated in a way that suppresses any taste or aversion for gambling on that mixture.

Sen rightly objects to requiring as a matter of rationality that choosing a lottery to determine which "pure" option is to be implemented be value neutral in this sense. He points out that there is nothing irrational about shunning a choice between "pure" options in S in favor of mixtures of such options so as to avoid direct responsibility for the element of S that is implemented. Dr. Chang must choose between giving the only dose of a given medicine available to one of two children in life or death circumstances where one of them has a slightly better chance of survival if treated than the other. He wishes to dodge making the choice and opts instead for the use of a random device even though the expected value of doing so is less than one of the pure options (175–76). There is nothing irrational in Chang's decision—although avoiding tough choices in this way may express value commitments rationalizing some form of moral cowardice.

---

[8] Sen seems prepared to recognize the possibility of converting menu dependent preference to menu independent preference by adjusting the universal set in some contexts (178). But he insists that when an agent follows a rule or mode of choice that is menu dependent, this approach is not allowed (178–79). However, I have just constructed a requisite universal set Ω′ for menu dependent rules that generates a menu independent preference and menu independent choice function. And I am arguing that the universal set constructed is appropriate for the evaluation of the rationality of such rule following choice. The menu dependence of the rule to be followed does not imply either menu dependent preference or choice function.

Where I disagree with Sen is in his contention that behavior like Chang's violates the independence postulate that enforces the value neutrality of mixtures. Chang's behavior is an indication that the lottery he chooses is not a member of the mixture set of neutral mixtures in $M(S)$. Instead of thinking of Chang's set S of pure options as the options of giving the treatment to one of the two candidate patients, the pure options belong to set T which is the set S plus the option of using the random device. The mixture set $M(T)$ is *not* a set of options available to Chang any more than the set $M(S)$ is. Chang can evaluate the options in $M(T)$ on the belief contravening supposition that they are available and on the additional belief contravening supposition of value neutrality. Notice that the option of using the random device to select a patient appears in the list twice: once as a pure option in T that is not a value neutral mixture and the other as a value neutral mixture. Under the suppositions as specified, the preference ranking over $M(T)$ obeys the independence postulate. The item that appears twice is described differently and assigned different values relative to the two descriptions. When described as a pure option of avoiding a direct decision as to who is to be treated, the value imputed is greater than the value of treating the patient most likely to be cured. Described as a value neutral mixed option, the value imputed is less than the value of treating the most promising patient. All of this conforms to the independence postulate and expected utility.

Such hypothetical reflection could bring clarity to Chang's deliberation. If Chang is conscientious, he might want to ponder whether the premium he is placing on avoiding the direct choice of the best pure option is worth the reduction in the expected benefit to the patient who would have otherwise received the treatment for sure. To ask this question presupposes the requirement of the independence postulate on the evaluation of value neutral mixtures as a condition on rational preference. But it does not mandate that rational agents evaluate the mixed options they face as options in a value neutral way.

Sen is aware of views of this kind. He declares that they rob expected utility theory of "much of its operational content" (176). The allegation may be true according to a very narrow reading of operational content. Choice functions characterize an agent's judgments as to what he would choose or judge admissible for choice in various hypothetical and, indeed, typically belief contravening circumstances. This is always the case. There is no diminution of operational content resulting from following the standard practice in Chang's case.

Nor does the account just proposed of the Dr. Chang example support his additional claim that the expected utility axioms are "trivially fulfilled" (176). Dr. Chang could evaluate a mixed option on the supposition that it is value neutral in just the same way he evaluates the lottery to decide which patient to treat. He would indeed violate the dictates of the independence postulates. Fulfilling the requirements of the axioms is not trivial. The issue is not whether Chang could violate conditions of rationality but whether he rationally should do so. Conformity with the independence postulate relative to a value neutral mixture set generated by the menu as a matter of rationality may, indeed, be trivial as a prescription. So is the requirement that a rational agent ought to be certain of the deductive consequences of the set of propositions of which he or she is certain. This form of triviality is a virtue in a prescriptive theory of rationality even though compliance with the prescriptive theory is far from trivial.

Sen is right to observe that expected utility principles are unsuccessful as predictive and descriptive models of human behavior in many circumstances. People lapse from standards of rationality. Moreover, as a prescriptive principle, the injunction to maximize expected utility is applicable only under conditions that depend on the contextual factors that are not always explicitly specified in the formula for expected utility. Sen acknowledges that expected utility theory is, when compared to alternatives, relatively successful (176, fn. 29). He seeks to identify its limitations. I agree that it has limitations both as descriptive theory and prescriptive principle. Chang's evaluation of the option of tossing the coin to determine which child is treated ought not depend exclusively on the expected utility calculation for coin tossing.[9]

---

[9] It may be worth mentioning that if Chang had at his disposal an alternative to a stochastic process for making the decision such as the advice of a seer or the dictates of the law, he might avoid responsibility for what he does in that way. These alternatives would also be rationally coherent even if morally questionable.

But this limitation undermines neither the expected utility principle nor the independence postulate.

In any case, whether we consider "choice consistency conditions" or "independence" postulates or other candidate requirements on rational choice, it is clear that requirements of rationality depend for their legitimacy not only on the menu of options available but on the decision maker's value commitments and beliefs. I have quibbled here with some of the respects in which Sen has brought this point to bear on various types of applications. But his diverse elaborations are unfailingly instructive and supportive of his general thesis that rationality is a discipline and neither a favored formula, nor an essentialist doctrine.

II. SELF-INTEREST MAXIMIZATION, THE PRISONER'S DILEMMA, AND THE ASSURANCE GAME

Sen has contended for a long time that the prisoner's dilemma and its more sophisticated variants can be understood as challenging the idea that rational agents should pursue their own goals "subject to feasibility considerations, without being restrained by any other values" (216). Sen makes clear that the goals challenged are not merely those restricted to promoting individual consumption without sympathy for others. Nor are they limited to efforts aimed at promoting one's own welfare where account is taken of the benefit accruing from sympathy with the benefits accruing to others (or the losses resulting from sympathy with the misery of others). The challenge Sen means to raise is against *self-goal choice.* "Each act of choice of a person is guided immediately by the pursuit of one's goal (and in particular, it is not restrained by the recognition of other people's pursuit of their goals)" (214).

Departing from self-goal choice is not to be confused with changing one's goals by incorporating into them consideration of the goals of others. To do the latter is changing one's goal. But the new goal, like the old one, is the agent's own goal. To be guided by that goal continues to be acting according to self-goal choice. Rejecting self-goal choice according to Sen appears to amount to this: even though the decision maker retains a given system of goals, the decision maker does not act in a manner that promotes those goals but rather in a fashion that "is constrained by the goals of others, or by rules of conduct, thereby violating self-goal choice (that is, the influences may affect the person's choice without their taking the form of goals that the person can be seen as pursuing himself)" (214).

In order to understand what Sen has in mind, recall the prisoner's dilemma. It is common wisdom that in the one shot prisoner's dilemma, rational players both defect because defecting is better than cooperating no matter what the other player does. Sen shares this common wisdom. Common wisdom is sensible if both players judge the other player's choice to be probabilistically independent of what he chooses. But player 1 may think it highly likely that player 2 will choose like he does. In that case, player 1 could find it better to cooperate than to defect. If player 2 did have similar kinds of beliefs to player 2, both players could perfectly rationally pursue their own goals and cooperate.

Sen is aware of this view but he does not seem to take it seriously himself. He thinks defection is rationally mandatory in the prisoner's dilemma provided that each player acts in a manner that is maximal given the agent's goals.

Sen seems to think that the players pay a price for self-goal choice. Both players acknowledge that they would be better off if they both cooperated.

I do not see that the players pay any price. Neither player has control over what the other player does. Each player's options are restricted to what he does. So even though both players recognize that both cooperating is pareto superior to both defecting, this fact is irrelevant since no agent has control over whether they jointly cooperate or jointly defect.

Sen thinks the lack of control can somehow be finessed. Sen reasons that "if the recognition that we can all better pursue our respective goals by jointly departing from goal priority [self-goal choice] makes us do exactly that, why should that departure change the nature of the goals that we are trying to pursue?" (212). Sen suggests that this sort of result can be brought off if both parties to a prisoner's dilemma acted *as if* the payoff structure of the game were different. In the payoff structure of the prisoner's dilemma, each player prefers his defecting and the other player cooperating to both cooperating and prefers both defecting to his cooperating and the other defecting. Sen suggests acting as if the payoff structure were that of the Assurance Game. Each player prefers both cooperating to his (or her) defecting and the other cooperating and prefers both defecting to his (or her) cooperating while the other defects. There are two equilibrium solutions to the Assurance Game game (both cooperating and both defecting).

The goals characterized by the payoffs in the Assurance Game do not represent the goals of either player. But suppose they act as if these were their goals. Moreover, suppose that the two players have enough confidence in each other to judge it likely that if their partner will cooperate, then they will both cooperate. Sen argues that both players would be better off according to their true goals acting as if their goals were those spelled out for the Assurance Game than they would be if they both acted by promoting the true goals. So they should not act in accordance with their true goals (those represented by the payoff structure of the prisoner's dilemma) but in accordance with the payoff structure of the Assurance Game.

Two objections come immediately to mind. The first is that the argument for acting in accordance with the payoff structure of the Assurance Game depends on each player having confidence that the other will cooperate—that is, on a prediction as to what the other player will do that cannot be derived solely from considerations of the other player's rationality and information about the predicament. Trust or confidence in the other is required as well. But as noted above, using the payoff structure of the prisoner's dilemma leads to cooperation if each player believes that the other player will choose in the way he does. So even if pretending that one's goals are different than they are were reasonable, it is not necessary once the centrality of confidence or trust in the other is appreciated.

Even if this point is waived, however, it is difficult to understand why the use of the payoff structure of the Assurance Game does not represent a change in each player's goals rather than an abandonment of self-goal choice. Perhaps prior to making the change, each player might reason that in the light of his current values and goals (for example, those represented by the payoff structure of the prisoner's dilemma) it would be best to alter his goals. This argument does not justify abandoning self-goal choice. In general, whether one is seeking to justify a change in beliefs or in values, the justification should be grounded in the beliefs and values prior to change. But once one implements the change one has concluded is to be made, the beliefs and values one endorses are different from what they were. But they are not "as if" beliefs and values.

In sum, I agree with Sen that rational agents can cooperate in prisoner's dilemmas—even in the absence of prior agreements and understandings. This view is in keeping with Sen's ecumenical attitude towards rationality. But I do not understand how a rational agent can do other than choose in a way that promotes the goals and values the agent endorses in the context of deliberation. Of course, rationality does not require that the goals and values be restricted to "self-centered welfare" or to what promotes a self-welfare goal maximizing the welfare of the agent even if the agent is benevolent and includes the interests of others in his or her welfare. It seems that Sen's tendency to think that we need to break not only with self-centered welfare or self-welfare but with self-goal choice derives from the failure to appreciate the possibility of changing one's goals and values for a reason while continuing to satisfy the requirements for self-goal choice.

A rational agent should choose an option that is optimal according to the agent's point of view if that point of view recognizes the availability of an optimal option. Other points of view do not count unless the agent makes those other points of view his own.

III. INCOMPLETE PREFERENCE

Deliberate (as opposed to routine or automatic) decision making requires that the decision maker ascertain enough information about his or her value commitments and beliefs as they relate to the menu of options S so as to identify a value structure $V(S)$ for S. Condition (ii) requires that the value structure $V(S)$ for menu S must be a weak order that guarantees that weak preference is reflexive, transitive, and complete, equipreference is reflexive, transitive, and symmetrical, strict preference is irreflexive, transitive, asymmetrical, and that strict preference and equipreference are connected: for every x and y in S xPy, xIy, or yPx.

Sen rightly challenges the thesis that rationality requires that $V(S)$ (or, indeed, $V(\Omega)$) should be a weak ordering.

> A chooser, who may have to balance conflicting considerations to arrive at a reflected judgment, may not, in many cases, be able to converge on a complete ordering when the point of decision comes. If there is no escape from choosing, a choice decision will have to be made even with incompleteness in ranking (160).[10]

Sen has been one of the rare economists (or philosophers for that matter) who has understood, rightly in my view, that "incompleteness in the ranking" may not be eliminated by the moment of choice.[11] The decision maker X may have to face a choice where, given the best information available to X at the time of choice, there is no optimal option—no option in the set of available alternatives at least as good as every other alternative. Sen correctly maintains that being confronted with such a predicament is not a mark of irrationality. To the contrary, it is a mark of honesty and good sense to refuse to "make up one's mind" when reflection on "all things considered" has not provided a warrant for doing so.

---

[10] In a footnote, Sen writes that incompleteness can arise from "limited information" or "unresolved conflict."

[11] Indeed, incompleteness in preference or evaluation is not widely studied by psychologists and social scientists interested in decision making. As a consequence, experimental data are misleadingly interpreted. For example, it is conventional wisdom that experimental data has conclusively established widespread violation of the independence postulate or sure thing principle. That is true only if the behavior observed is taken to reveal definite preferences rather than choice in the face of unresolved conflict.

In this connection, Sen draws an important distinction between "tentative incompleteness" and "assertive incompleteness." The decision maker faces a decision problem with tentative incompleteness in his value structure if the agent lacks the opportunity to pursue matters further and must make a decision even though he or she thinks that given sufficient time, opportunity, and resources, a warranted resolution removing the incompleteness would be forthcoming. Assertive incompleteness obtains if the decision maker despairs of the possibility of resolving the conflicts in his or her value commitments even "in principle." Sen wishes to provide room for the sort of tragic choices that Bernard Williams, Martha Nussbaum, and Charles Larmore and other acolytes of Isaiah Berlin insist upon. I agree that principles of rationality ought not to preclude such predicaments. But as a matter of methodological morals, it seems to me that we should avoid placing roadblocks in the path of inquiry as Peirce would say. No unsettled issue either in science, morals, or politics should be taken to be irresolvable in principle unless we have an impossibility theorem. Otherwise we should continue to entertain the "hope" as Peirce would say that future inquiry could bring resolution for similar kinds of issues.[12]

Sen's distinction and the issues it raises are philosophically important. They go to the heart of the question as to the extent to which problems in ethics are amenable to inquiry. But as far as the question of rational choice is concerned, the topic may be bypassed as Sen clearly recognizes. Whether incompleteness at the moment of choice is tentative and the conflict in values can in principle be resolved or is assertive in a way that precludes resolution, the exigencies of choice require deciding without resolving the conflict. In order to provide a framework for addressing such predicaments, the account of value structure for the menu of options given in condition (ii) needs to be amended. The requirement of completeness must be removed from the condition of a weak ordering.

---

[12] See Levi, "Conflict and Inquiry," *Ethics*, cii (1992): 814–34, reprinted in *The Covenant of Reason* (New York. Cambridge, 1997), chapter 11. It is important not to confuse Sen's distinction between tentative incompleteness and assertive incompleteness with another important contrast between imprecision and indeterminacy. Agent X may be committed to a value structure $V(S)$ representable by a weak ordering over S. Yet X may not be able to identify fully that weak ordering with complete precision. X has failed to fulfill X's commitments There is incompleteness in X's "performance." At the moment, X cannot explicitly indicate whether X prefers a over b, b over a, or equiprefers a and b Yet, there is neither tentative nor assertive incompleteness in X's value structure. By way of contrast, Y might be able to identify Y's value structure as one that is incomplete. We may say that X is ignorant of the details of X's value structure. X is in need of therapy, training, or aid in computation to fulfill X's commitment. Y is fully aware of the incompleteness in Y's value structure. There is no need for therapy, training, and so forth. *Inquiry* will, if successful, remove tentative incompleteness. Neither therapy nor inquiry can remove assertive incompleteness. Needless to say, agent Z might be committed to an incomplete value structure and not be able to identify it precisely

What, however, is entailed by giving up completeness? Recall that for Sen and many others, a complete ordering is one representable by a weak preference relation R that is reflexive, transitive, and complete and where strict preference is the asymmetric factor of R in the sense that xPy if and only if $xRy\&\sim yRx$ while equipreference is the symmetric factor so that xIy if and only if $Xry\&yRx$. Sen calls the weak preference with reflexivity and transitive but without completeness a *quasi ordering*. The factorization of weak preference into strict preference and equipreference is taken to be a matter of definition so that when completeness of weak preference is given up, Sen retains the definitions of strict preference and indifference. However, Sen's practice here is very much open to question for the same reasons that justify calling the requirement of completeness into question.

Incompleteness in preference arises when the value commitments, goals, and values of the decision maker impose constraints on the way that the decision maker ranks the options available in S that allows several different rankings to be permissible. X has promised to pay Y a certain sum of money by a certain date in return for specified services. Y renders the services and X confronts the decision whether to pay the sum in cash or by check or to refuse to pay the bill. X acknowledges the obligation to keep promises and, hence, ranks refusal to pay below the other alternatives. How should X rank the other alternatives? An obligation to keep promises may not impose any constraint on how X ranks these options. Three rankings are permissible:

> *Way 1*: Paying by check and by cash are equipreferred (and both preferred to refusal to pay).
>
> *Way 2*: Payment by check might be ranked above payment in cash and paying by cash over nonpayment.
>
> *Way 3*: Payment in cash may be ranked over payment by check that is again ranked over nonpayment.

The obligation to keep promises and pay debts rules out none of these rankings. If X's value commitments impose no additional constraints, X is committed to recognize all three rankings as permissible rankings of these options. Of course, *X* could endorse further value commitments that rule out one or more of these permissible rankings as impermissible. Ways 2 and 3 might be ruled out because *X* judges it a matter of indifference whether he pays by check or by cash. *X* might wish to help *Y* evade taxes and rule out ways 1 and 2. *X* might simply think that it is not his obligation to impede *Y*'s tax evasion so that *X* rules 2 but neither 1 nor 3. *X* could see *X*'s duty to be to impede such tax evasion and rule out ways 1 and 3. Or *X* might think that it is not his obligation to impede such tax evasion and thus rule out way 3 alone.

X cannot, however, reasonably rule out 1 while retaining 2 and 3 as permissible. If *X* is in doubt as to whether paying by cash is better than paying by check or the other way around, then ranking the two options together should be permissible as well. The latter ranking is a "potential compromise" between the other two and should be regarded as permissible if they are.

In any case, if all three ways of evaluating the options are recognized to be permissible, the only constraint *X* endorses on *X*'s evaluations is the obligation to keep *X*'s promises. *X* clearly recognizes that both paying in cash and by check are strictly better than refusing to pay. Every permissible ranking agrees on that point. But paying by cash and paying by check are noncomparable. Neither option is weakly preferred to the other. And no option is weakly preferred to every other option. According to standard definitions of optimality including Sen's, there is no optimal option among the three available. *X* cannot optimize even though *X* is rational.

Now the notion of preference on which I relied to describe this illustration characterizes a value structure in terms of the weak orderings or rankings of the options recognized as permissible according to the constraints generated by *X*'s value commitments. Option x is categorically weakly preferred to option y if and only if x is weakly preferred to y according to all permissible rankings in the value structure. x is categorically strictly preferred to y if and only if x is strictly preferred to y according to every permissible ranking and is categorically equipreferred to y if and only if x is equipreferred to y according to every permissible ranking.

The notion of categorical weak preference deployed here is formally similar to Sen's notion of a binary relation of aggregation R^* in chapter 7* of *Collective Choice and Social Welfare* and the notion of a weak preference derived by "utility based intersection" discussed on pp. 362–63 of chapter 11 of the book under review. In developing a basis for classifying different kinds of interpersonal comparisons, Sen considers different rankings of the social states or alternatives derived from different ways of "aggregating" the preferences or valuations of the individual citizens or beneficiaries of social welfare. The details do not matter here. It is sufficient to notice that a given family of such rankings in the setting of social choice corresponds to a family of permissible rankings in the value structure of an individual decision maker. The weak preference relation R^* is defined in a fashion entirely analogous to the way I defined categorical weak preference.

There is, however, one important difference between how Sen thinks of incompleteness in valuing or preference and incompleteness in categorical preference. According to Sen, strict preference and equipreference are obtained by factoring weak preference into its asymmetric and symmetric parts. Categorical preference as I have defined it is not always factorizable in this fashion.

Return to our example. Suppose X rules way 3 out as impermissible. The other two ways are recognized as permissible. In that case, paying by check is categorically weakly preferred to paying by cash since it is weakly preferred according to every permissible ranking. However, it is neither categorically equipreferred nor categorically strictly preferred to paying by cash.

However, it is also true that paying by cash is not categorically weakly preferred to paying by check. According to Sen's approach, paying by check is strictly preferred to paying by cash. Whereas I derive the claim that X does not categorically strictly prefer paying by check over paying by cash from the given value structure and categorical weak preference, Sen derives exactly the opposite from the *same* value structure and categorical weak preference. In effect, Sen's approach reduces the value structure that recognizes ways 1 and 2 as permissible and 3 as impermissible to a value structure where only 2 is recognized to be permissible. Sen does this by resorting to the standard practice of defining strict preference as the asymmetric component of weak preference.

According to both approaches, the categorical weak preference is a quasi ordering. But Sen's approach sees the quasi ordering as free of any indeterminacy. Strict preference and equipreference are connected. Given any pair of options in the triple, either one is strictly preferred to the other or they are equipreferred. In facing a three way choice, Sen's view recognizes paying by check as uniquely optimal and maximal. An option is maximal if there is no option in the menu of available options strictly preferred to it. This is Sen's favored criterion for choice when there is incompleteness in the value structure. Paying by check is strictly preferred to paying by cash according to Sen's understanding of strict preference. According to Sen's recommended approach, paying by cash is not admissible for choice.

According to the approach I use, paying by check is uniquely optimal but it is not uniquely maximal. Paying by check is not categorically strictly preferred to paying by cash. Maximality, Sen's recommended criterion for choice, allows for the choice of either option.

Thus, the difference here does not concern how one should make choices in the face of incompleteness in one's valuations or preferences. I am assuming Sen's principle of maximality. The problem arises because Sen understands the relation of aggregation or categorical preference as taking strict preference to be the asymmetric factor of weak preference. The difference between Sen's approach and the one I favor is no issue of terminological legerdemain. Sen's practice rules out the relevance of equipreference between a pair of options *a* and *b* as a permissible ranking when there is a permissible strict preference for *a* over *b* but not the other way around. Sen is prepared to recognize certain kinds of incompleteness in value structure. And his championing this view and appreciation of its significance is a first rate contribution. My reservation with his account concerns one of those rare cases where Sen has failed to be ecumenical enough in recognizing as rational certain kinds of value commitments.

Given Sen's own ambition to take a generous, comprehensive view of rationality, it seems to me that when he rightly acknowledges the rationality of having incomplete preferences, he should also recognize the rationality of sometimes having complete but disconnected preferences as well. As matters stand, Sen denies the rational coherence of the latter idea.

IV. MAXIMALITY VERSUS V-ADMISSIBILITY

The value structures used to derive the weak preference quasi orderings to which maximality applies can themselves be derived from sets of expected utility functions in ways that need not detain us here but that provide a way of accommodating decision making under uncertainty within the purview of the current discussion. In this setting, maximality as a criterion of rational choice has the virtue of, on the one hand, allowing for incomplete preferences and complete but disconnected categorical preferences while, on the other hand, prohibiting, in agreement with the classical expected utility theory, choosing options that are "dominated" by other options.

But maximality is not the only candidate criterion of rational choice that exhibits this double-barreled virtue.

Call an option *V*-admissible in a menu set *S* if and only if there is a permissible utility function in the value structure *V*(*S*) according to which the option is ranked as optimal. All *V*-admissible options must be maximal. There can be no option in *S* that is categorically strictly preferred in S to an option that is V-admissible in S. But there can be maximal options in S that are not V-admissible. An abstract example will have to suffice here. Imagine a choice between three options x, y, and z. According to one permissible utility function in $V(S)$, $v(x) = 1$, $v(y) = 0.4$, and $v(z) = 0$. According to another, $v(z) = 1$, $v(y) = 0.4$, and $v(x) = 0$. The set of permissible utility functions is the set of all weighted averages of these two (and positive affine transformations of these). y cannot come out optimal according to any such utility function. Only x and z can. So x and z are V-admissible but y is not. Yet y, along with x and z are maximal.

So we have two candidate criteria for choice that allow incompleteness in ordering options in S and respect the independence postulate. Is there any basis for deciding between them?

In the ecumenical spirit that is exhibited by Sen's attitude towards rationality, one could argue in favor of maximality on the grounds that it allows more options than V-admissibility does to be admissible given the value commitments and beliefs of the decision maker.

But the *same* ecumenical spirit should allow us to recognize an important difference between the value structure for the example just given and a variant of it where $v(y)$ is given 0.6 in both rankings. We should allow rational agents to be sensitive to such differences between second worst and second best cases. Maximality does not do well by this standard. It *forbids* taking such differences to be relevant in decision making. V-admissibility, on the other hand, has resources sufficient to distinguish between agents whose value commitments do not recognize a relevant distinct between second worst and second best cases and agents whose value commitments entail such recognition. V-admissibility countenances the rationality of both kinds of agents.

V. CHOICE CONSISTENCY AGAIN

Throughout this discussion, I have been assuming that whether or not value structures for a universal set Ω and nonempty finite subsets of Ω exhibit preference independence and that whether the choice functions are derived using maximality or V-admissibility, they will be menu independent as well.

Under this assumption, maximality, as Sen acknowledges, will violate β. It will, however, conform to α and γ. Given a class M of option or menu sets, x is in the maximal set for every S in M if and only if it is in the maximal set for the union of all sets in M. When preference is menu independent, Sen is willing to acknowledge that rational agents ought to conform to α and γ.

Advocates of V-admissibility are in this respect harsher critics of choice consistency requirements than Sen appears to be. Even when preference is menu independent, choice functions defined by V-admissibility violate γ.[13] This is another respect in which V-admissibility is more congenial with a comprehensive, ecumenical approach to rationality than maximality.

VII. CONCLUSION

As I have already indicated, Sen's book focuses as explicitly as many of Sen's books do on devising ways and means of adapting the analytic devices favored by economists to furnish a way to incorporate a serious recognition of moral and political values within the framework of economic analysis—especially social choice theory. As I indicated at the outset, I have not touched at all on the rich treatment of how considerations of freedom may be accommodated within the framework of social choice theory. And only the surface of Sen's approach to developing a comprehensive ecumenical approach to rationality has been scratched.

I have argued that although Sen's attack on excessively formulaic requirements on rationality such as internal choice consistency is welcome and the idea of menu dependent preference is an important topic for critical scrutiny, more motivated accounts of choice consistency based on maximizing in Sen's sense in a setting of menu independent preference insist on requirements for choice consistency that are stronger than what rationality should require.

The cogency of this complaint depends in large measure on the contention that maximality is not comprehensive and ecumenical enough as a principle of rational choice when preferences are incomplete.

Having said this, it must be emphasized that the dispute between advocates of maximality such as Hans G. Herzberger, Sen, and Peter Walley[14] and those sympathetic with V-admissibility is a controversy among allies sharing a common preoccupation with forging a framework for normative principles of rationality tolerant of diversity while insistent on a respect for logicality and reason. Sen's book is testimony to his abiding, serious commitment to and insightful and rigorous pursuit of this vision.

ISAAC LEVI

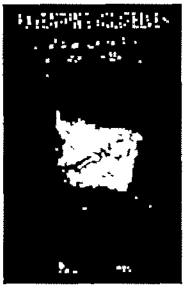
Columbia University

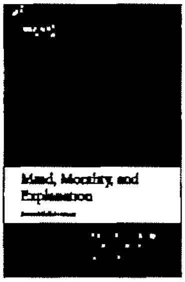
---

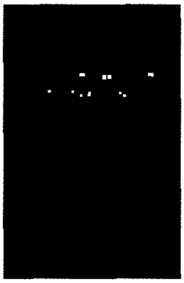
[13] Levi, *Hard Choices.*

[14] In addition to Sen, Herzberger, "Ordinal Preference and Rational Choice," *Econometrica*, XLI, 2 (March 1973): 198–99; and Walley, *Statistical Reasoning with Imprecise Probabilities* (London: Chapman and Hall, 1991), pp 162–66.

New from
Oxford







Prices are subject to change and apply only in the US. To order, or for more information, please call 1-800-451-7556. Visit our website at www.oup.com/us

OXFORD

Extending Ourselves
Computational Science, Empiricism, and Scientific Method
PAUL HUMPHREYS
Computational methods have become the dominant technique in many areas of science. This book contains the first systematic philosophical account of these new methods and their consequences for scientific method.
June 2004 $45.00

Mind, Morality, and Explanation
Selected Collaborations
FRANK JACKSON, PHILIP PETTIT, and MICHAEL SMITH
Frank Jackson, Philip Pettit, and Michael Smith have been at the forefront of philosophy in Australia for much of the last two decades, and their collaborative work has had widespread influence throughout the world. *Mind, Morality, and Explanation* collects the best of that work in a single volume, showcasing their seminal contributions to philosophical psychology, the theory of psychological and social explanation, moral theory, and moral psychology.
May 2004 paper $29.95 cloth $74.00

Ancient Worlds, Modern Reflections
Philosophical Perspectives on Greek and Chinese Science and Culture
GEOFFREY LLOYD
Geoffrey Lloyd's pioneering new book uses a study of ancient Greek and Chinese science and culture to throw light on fundamental problems, both intellectual and moral, that we still face today. The issues range from the debate about realism and relativism in philosophy of science to doubts concerning the universal applicability of the discourse of human rights.
April 2004 $35.00

Consciousness and Its Objects
COLIN MCGINN
Colin McGinn presents his latest work on consciousness in ten interlinked papers, four of them previously unpublished. He extends and deepens his controversial solution to the mind-body problem, defending the view that consciousness is both ontologically unproblematic and epistemologically impenetrable. He also investigates the basis of our knowledge that there is a mind-body problem, and the bearing of this on attempted solutions.
May 2004 $35.00

The Provenance of Pure Reason
Essays in the Philosophy of Mathematics and Its History
WILLIAM TAIT
This volume collects William Tait's most important published philosophical papers from the 1980's to the present. The articles cover a wide range of issues in the foundations and philosophy of mathematics, including some on historical figures ranging from Plato to Gödel.
May 2004 $60.00

New from FORDHAM

Debates in Continental Philosophy
Conversations with Contemporary Thinkers
Richard Kearney
388 pages
ISBN 0-8232-2317-5, Hardcover, $65 00
ISBN 0-8232-2318-3, Paperback, $24 00
Perspectives in Continental Philosophy, No 37

Thinking Difference
Critics in Conversation
Julian Wolfreys
256 pages
ISBN 0-8232-2307-8, Hardcover, $55 00
ISBN 0-8232-2308-6, Paper, $22 00
Perspectives in Continental Philosophy, No 35

Being Jewish/Reading Heidegger
An Ontological Encounter
Allen Scult
186 pages
ISBN 0-8232-2311-6, Hardcover, $35 00
Perspectives in Continental Philosophy, No 36

Hans Urs von Balthasar and the Dramatic Structure of Truth
A Philosophical Investigation
D C Schindler
640 pages
ISBN 0-8232-2321-3, Hardcover, $65 00
Perspectives in Continental Philosophy, No 34

Quantum Mechanics and the Philosophy of Alfred North Whitehead
Michael Epperson
256 pages, illustrated
ISBN 0-8232-2319-1, Hardcover, $55 00
American Philosophy, No 14

Conversion in American Philosophy
Exploring the Practice of Transformation
Roger Ward
320 pages
ISBN 0-8232-2313-2, Hardcover, $70 00
American Philosophy, No 13

Fordham University Press
[illegible]

Essential.

The Lesser Evil

Political Ethics in an Age of Terror

Michael Ignatieff

There is perhaps no greater political challenge today than trying to win the war against terror without losing our democratic souls.

Michael Ignatieff meets this challenge head-on, with the combination of pragmatic idealism, historical awareness, and political judgment that has made him a widely respected writer and commentator. He stakes out a new, and philosophically informed, position: in favor of force to support democracy and thoughtful restraint in its application.

Cloth $21.95 ISBN 0-691-11751-9

The Good in the Right

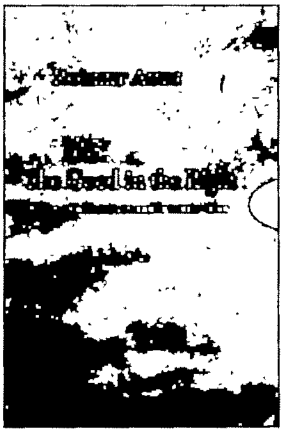
A Theory of Intuition and Intrinsic Value

Robert Audi

This book represents the most comprehensive account to date of an important but widely contested approach to ethics—intuitionism, the view that there is a plurality of moral principles, each of which we can know directly. Robert Audi casts intuitionism in a form that provides a major alternative to the more familiar ethical perspectives.

"*The Good in the Right* will end up being thought of as one of the most important books in moral philosophy published in this decade."
—Brad Hooker, author of *Ideal Code, Real World*

Cloth $35.00 ISBN 0-691-11434-X



PRINCETON
University Press

800-777-4726 • READ EXCERPTS ONLINE
WWW.PUP.PRINCETON.EDU

New books from The MIT Press

The Mind Incarnate

Lawrence A. Shapiro

"This book should lead scientists and philosophers to reconsider the pervasive assumption that the mind can be understood without reference to the material basis of mental processes." — Elliott Sober, Stanford University

Life and Mind series • A Bradford Book
272 pp., 23 illus. $34

Anti-Individualism and Knowledge

Jessica Brown

"Will be a milestone in the understanding of anti-individualism. . . . It sets a clear agenda for the next stage of the debate." — Paul Snowdon, University College London

Contemporary Philosophical Monographs series
A Bradford Book • 360 pp. $25 paper

Causation and Counterfactuals

edited by John Collins, Ned Hall, and L. A. Paul

A collection of important recent work on the counterfactual analysis of causation.

Representation and Mind series
A Bradford Book • 480 pp., 54 illus. $40 paper

Freedom and Determinism

edited by Joseph Keim Campbell, Michael O'Rourke, and David Shier

A state-of-the-art collection of previously unpublished essays on the topics of determinism, free will, moral responsibility, and action theory, written by some of the most important figures in these fields of study.

Topics in Contemporary Philosophy
A Bradford Book • 352 pp., 8 illus. $35 paper

Natural Minds

Thomas W. Polger

"Strongly recommended to anyone with a serious interest in the metaphysics of the mind." — Jaegwon Kim, Brown University

A Bradford Book • 320 pp., 15 illus. $38

Naturalistic Realism and the Antirealist Challenge

Drew Khlentzos

"I know of no other book in recent years that can serve, as this book can, at once as a comprehensive survey of one of the most important debates in contemporary metaphysics and a genuinely important contribution to that debate." — Hilary Putnam, Harvard University

Representation and the Mind series
A Bradford Book • 448 pp., 10 illus. $40

Organisms and Artifacts

Design in Nature and Elsewhere

Tim Lewens

"This book will give an old problem really new life, and must be the starting point for all future discussion." — Michael Ruse, Florida State University

Life and Mind series • A Bradford Book
240 pp., 5 illus. $32

Real Natures and Familiar Objects

Crawford L. Elder

"This book will be essential reading for philosophers who address questions concerning the nature of folk objects, the basis for modal claims and our knowledge of such properties, causation, vagueness, and much else." — William R. Carter, North Carolina State University

A Bradford Book • 224 pp. $32

Critical Resistance

From Poststructuralism to Post-Critique

David Couzens Hoy

"There is no better guide through the thickets of poststructualism and its aftermath." — Thomas McCarthy, Northwestern University

288 pp. $35

Adorno's Negative Dialectic

Philosophy and the Possibility of Critical Rationality

Brian O'Connor

"O'Connor's study makes Adorno's vital and detailed contributions to epistemology and metaphysics harder than ever to ignore." — Simon Jarvis, University of Cambridge, author of *Adorno: A Critical Introduction*

Studies in Contemporary German Social Thought
224 pp. $35

Case Studies in Biomedical Research Ethics

Timothy F. Murphy

"A thought-provoking collection of cases in biomedical research ethics that is astonishing in its breadth."
— Mark H. Waymack, Loyola University Chicago

Basic Bioethics series • 368 pp. $29 paper

now in paperback

What Genes Can't Do

Lenny Moss

"No other book presents such a comprehensive history and critique of modern biological thought."
— Robert Perlman, University of Chicago, editor of *Perspectives in Biology and Medicine*

Basic Bioethics series
256 pp., 2 illus. $20 paper

To order call 800-405-1619
Prices subject to change without notice

http://mitpress.mit.edu

Three Philosophical Filmmakers

Hitchcock, Welles, Renoir

Irving Singer

"Singer's book provides supremely literate commentary: this is a film criticism of ideas, images, and detail."
— James Engell, Harvard University

312 pp., 4 illus. $34.95

Writing the Future

Progress and Evolution

edited by David Rothenberg and Wandee J. Pryor

"An interesting and highly leavened mix of thought, perception, and speculation, all of which provokes more of the same." — David Appelbaum, SUNY at New Paltz

280 pp., 17 illus. $29.95

Zone Books

In Praise of Blandness

Proceeding from Chinese Thought and Aesthetics

François Jullien
translated by Paula M. Varsano

A consideration of blandness not as the absence of defining qualities but as the harmonious union of all potential values—an infinite opening into human experience.

Distributed for Zone Books
152 pp., 8 illus. $28

now in paperback

Detour and Access

Strategies of Meaning in China and Greece

François Jullien
translated by Sophie Hawkes

An exploration of the central role of indirect modes of expression in ancient China.

Distributed for Zone Books
435 pp. $22 paper

CAMBRIDGE

OUTSTANDING SCHOLARSHIP FROM CAMBRIDGE

Truth and Truthmakers
D.M. Armstrong
Cambridge Studies in Philosophy
$55.00* Hardback 0-521-83832-0 c 180pp
$21 00* Paperback 0-521-54723-7

New Essays on the History of Autonomy
A Collection Honoring J.B. Schneewind
Edited by Natalie Brender and Larry Krasnoff
$70 00 Hardback 0-521-82835-X 224pp

Satisficing and Maximizing
Moral Theorists on Practical Reason
Edited by Michael Byron
$65 00* Hardback 0-521-81149-X c 264pp
$23 00* Paperback 0-521-01005-5

The Ethics of Assistance
Morality and the Distant Needy
Edited by Deen K. Chatterjee
Cambridge Studies in Philosophy and Public Policy
$65 00 Hardback 0-521-82042-1 304pp
$24 00 Paperback 0-521-52742-2

Durkheim's Philosophy Lectures
Notes from the Lycée de Sens Course, 1883-1884
Edited by Neil Gross and Robert Alun Jones
$60 00* Hardback 0-521-63066-5 c 336pp

Information and Meaning in Evolutionary Processes
William F. Harms
Cambridge Studies in Philosophy and Biology
$65 00 Hardback 0-521-81514-2 272pp

Subjective Probability
The Real Thing
Richard Jeffrey
$60 00 Hardback 0-521-82971-2 144pp
$22 00 Paperback 0-521-53668-5

Relevant Logic
A Philosophical Interpretation
Edwin D. Mares
$65 00 Hardback 0-521-82923-2 240pp

Morality and Politics
Edited by Ellen Frankel Paul, Fred D. Miller, Jr., and Jeffrey Paul
Social Philosophy and Policy
$25 00 Paperback 0-521-54221-9 374pp

Leibniz and China
A Commerce of Light
Franklin Perkins
$65 00 Hardback 0-521-83024-9 242pp

Embryology, Epigenesis and Evolution
Taking Development Seriously
Jason Scott Robert
Cambridge Studies in Philosophy and Biology
$60.00 Hardback 0-521-82467-2 174pp

Understanding Phenomenal Consciousness
William S. Robinson
Cambridge Studies in Philosophy
$65 00 Hardback 0-521-83463-5 280pp

Mind World
Essays in Phenomenology and Ontology
David Smith
$75 00 Hardback 0-521-83203-9 328pp
$28 00 Paperback 0-521-53973-0

Popper, Otto Selz and the Rise of Evolutionary Epistemology
Michel ter Hark
$65 00 Hardback 0-521-83074-5 262pp

Our Knowledge of the Past
A Philosophy of Historiography
Aviezer Tucker
$70 00 Hardback 0-521-83415-5 304pp

*Prices subject to change

www.cambridge.org CAMBRIDGE UNIVERSITY PRESS

Calcutta University
ALIPORE CAMPUS

THE JOURNAL OF
PHILOSOPHY

[illegible]

VOLUME CI, NUMBER 6
JUNE 2004

page

277 *The Ubiquitous Problem of Empty Names* Stuart Brock

299 *Dialectical Delicacies in the Debate about Freedom and Alternative Possibilities* Ishtiyaque Haji and Michael McKenna

COMMENTS AND CRITICISM

315 *Stopping to Reflect* M.J. Schervish, T. Seidenfeld, and J.B. Kadane

323 NEW BOOKS

Published by The Journal of Philosophy, Inc.

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Akeel Bilgrami, Arthur C. Danto, Kent Greenawalt, Patricia Kitcher, Philip Kitcher, Isaac Levi, Mary Mothersill, Philip Pettit, Carol Rovane, Achille C. Varzi. *Editor Emeritus*: Sidney Morgenbesser. *Consulting Editors*: David Albert, John Collins, James T. Higginbotham, Charles D. Parsons, Wilfried Sieg. *Managing Editor*: John Smylie.

THE JOURNAL OF PHILOSOPHY is owned and published by the Journal of Philosophy, Inc. *President*, Arthur C. Danto; *Vice President*, Akeel Bilgrami; *Secretary*, Daniel Shapiro; *Treasurer*, Barbara Gimbel; *Other Trustees*: Lee Bollinger, Leigh S. Cauman, Kent Greenawalt, Michael J. Mooney, Lynn Nesbit.

All communications to the Editors and Trustees and all manuscripts may be sent to John Smylie, Managing Editor, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, New York 10027. FAX: (212) 932-3721. You may also visit our website at: www.journalofphilosophy.org

THE JOURNAL OF PHILOSOPHY

2004

SUBSCRIPTIONS (12 issues)

Individuals	$35.00
Libraries and Institutions	$75.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$15.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are required.

Published monthly as of January 1977; typeset and printed by Capital City Press, Montpelier, VT.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, if the copier abides by the JOURNAL's terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

POSTMASTER: Periodical postage paid at New York, NY, and other mailing offices.
POSTMASTER: Send address changes to the *Journal of Philosophy* at MC 4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027.

© Copyright 2004 by the Journal of Philosophy, Inc. ISSN 0022-362X

THE JOURNAL OF PHILOSOPHY
VOLUME CI, NO. 6, JUNE 2004

THE UBIQUITOUS PROBLEM OF EMPTY NAMES*

Until a little under thirty years ago, the orthodox view about proper names was that they had a meaning (or sense), that this meaning could be captured by a description of some kind or other, and that these names managed to refer to a particular individual because the descriptions associated with the names were definite descriptions. Something like this view was originally and independently advocated by Russell and Frege, and championed more recently by philosophers such as John Searle and Michael Dummett.[1]

---

* I am grateful for discussion and comments on earlier drafts of this paper from Nicholas Agar, Alan Baker, Paul Benacerraf, John Bigelow, Tony Fielding, Lloyd Humberstone, Frank Jackson, Seahwa Kim, Edwin Mares, Michael Nelson, Gideon Rosen, Scott Soames, and especially David Lewis.

[1] Of course, this is not a complete specification of the descriptive theorist's position. She will have to say much more before she can claim to be proffering a theory about the semantics of proper names. The descriptivist will at least have to provide an answer to the following questions:

(i) What kind of association is to hold between the name and the description? Can the name-description pair vary from speaker to speaker? Are they occasion dependent, so that even for a single speaker, there may be different name-description pairs on different occasions?

(ii) How close does the fit between an individual and set of descriptions have to be before the name associated with the descriptions is to count as referring to that individual?

(iii) Must the relevant descriptions associated with a name pick out a unique individual? What happens if there is more than one candidate for the referent of the name?

(iv) How are we to analyze the relevant definite descriptions? Should we follow Frege and treat them as singular terms ("Über Sinn und Bedeutung," *Zeitschrift für Philosophie und Philosophische Kritik*, C (1892): 25–50, "On Sense and Meaning," Max Black, trans., *Philosophical Review*, LVII (1948): 209–30)? Should we follow Russell and treat them quantificationally ("On Denoting," *Mind*, XIV (1905): 479–93)? Should we follow Delia Graff and treat them (at least sometimes) as predicates ("Descriptions as Predicates," *Philosophical Studies*, CII (2001): 1–42)?

(v) What kind of description will count as a reasonable candidate for capturing the sense or meaning of a proper name? Saul Kripke has shown us that not just any uniquely identifying description will do.

0022-362X/04/0106/277–98

Descriptive theories have recently been contrasted with direct reference theories. The most basic tenet of any direct reference theory is that the semantic value of any proper name is nothing over and above the individual to which it refers. According to this view, the meaning of a name cannot adequately be captured by a description of any kind. Direct reference theorists typically supplement this tenet with the further thesis that any simple sentence containing an occurrence of a proper name expresses a singular proposition—a proposition in which the individuals referred to occur as constituents—if it expresses a proposition at all. This view was originally advocated by Mill, has been vigorously defended by authors such as David Kaplan, Nathan Salmon, and Scott Soames, and has been supported by arguments found in the work of Keith Donnellan, Saul Kripke, Ruth Barcan Marcus, and Hilary Putnam. Indeed, it would not be unreasonable to suggest that this view has become the new orthodoxy.

When philosophers do their accounting—that is, when they come to compare the merits and shortcomings of the two theories—it is often suggested that sentences containing an occurrence of an empty name are not well accounted for by the direct reference theorist, whereas descriptive theories can handle them with ease. Such a pronouncement has been made by descriptivists and direct reference theorists alike.

Descriptivists have often taken such considerations as a significant point against—if not a complete refutation of—direct reference theories, without even conceiving that similar considerations might also tell against their particular variety of descriptivism. To illustrate, consider the following:

> Russell instructs us to test a logical and semantical theory by "its capacity for dealing with puzzles." His own theory of ordinary proper names nicely passes muster with respect to...[two] such puzzles: those presented by empty proper names, [and, as a special case] by negative existentials containing proper names.... On the other hand the anti-Fregean view... founders on these rocks.[2]

> [The Kripkean] criticism of the description theory has been widely accepted as showing that no description theory of proper names is tenable. As a consequence...the problem that Millians face in the area of language has gone virtually unnoticed. This is unfortunate. The cause of philosophical understanding as well as the cause of descriptivism would have been better served had the problem received the attention

---

[2] Alvin Plantinga, "The Boethian Compromise," *American Philosophical Quarterly*, xv (1978): 129–38, here p. 129–30.

> it deserves...[for] in the area of language, a strong case can be made against Millianism, in connection with the phenomena of names without bearers.[3]

Likewise, direct reference theorists tend to acknowledge that empty names present a problem for their theory. But when they consider the problem directly, direct reference theorists either attempt to explain away the problem or point to more general issues in the philosophy of language which are supposed, on balance, to vindicate their semantic theory. The interesting point to notice when one comes across any such evaluation, however, is that it is almost universally conceded—either implicitly or explicitly—that the descriptivist has no problem on this score. Consider the following appraisal of the philosophical landscape set forth by three direct reference theorists:

> There are two major types of theories of the meaning of proper names: Millian and Fregean.... Each theory (type) has its apparent successes and failures, and, often, what is regarded as one's failure is viewed as its rival's success.... [There are] two such problems for the Millian theory...the problem of negative existentials...and the problem of non-referring names.[4]
>
> The issues surrounding the use of non-denoting names are of *particular* interest to those philosophers who hold to a version of the direct reference theory for names and indexicals...Russell solves the problem of non-denoting proper names by claiming that proper names should be semantically viewed as definite descriptions.... From the direct reference perspective, however, Russell's puzzle about non-denoting proper names remains a puzzle.[5]
>
> In "On Denoting," Russell trumpeted his Theory of Descriptions...for its ability to handle a variety of puzzles that arise on his theory that the semantic content of a singular term is solely its referent (denotation, designatum). The puzzles are primarily: Frege's Puzzle about $\ulcorner \alpha = \beta \urcorner$; the more general problem of substitution failure in certain contexts, especially those ascribing propositional attitude; the question of content and truth value for sentences involving nonreferring terms; and as a special case, true negative existentials.... It has been objected that the

---

[3] Jerrold Katz, "Names without Bearers," *Philosophical Review*, CIII (1994): 1–39, here pp. 2–3.

[4] T.C. Ryckman, "The Millian Theory of Names and the Problem of Negative Existentials and Non-referring Names," in D.F. Austin, ed., *Philosophical Analysis: A Defense by Example* (Boston: Kluwer, 1988), pp. 241–50, here p. 241.

[5] Gregory Fitch, "Non Denoting," *Philosophical Perspectives*, VII (1993): 461–86, here pp. 461–62.

> second two problems are sufficient by themselves to refute Millianism even if the first two problems are not.[6]

Salmon later essentially concedes that the problem is a special problem for direct reference theorists:

> A Millian like myself...may not avail him/herself of the Theory of Descriptions to solve the problem of sentences with non-referring names. If α is a proper name, referring or not, it is not a definite description, nor by the direct reference theorist's lights does it 'abbreviate' any definite description. Direct reference theory thus excludes application of the Theory of Descriptions [to solve this problem].... For similar reasons, the direct reference theorist is also barred from using Frege's sense-reference distinction to solve the difficulties (*ibid.*, p. 285).

It is my contention, however, that this common perspective is mistaken. To the extent that the direct reference theorist has a problem with empty names, the descriptivist has one also, and it is the aim of this paper to explain why.

I. DIRECT REFERENCE (AND THE PROBLEM OF EMPTY NAMES)

In order to illustrate the problem empty names present the direct reference theorist, consider the following sentences.

(1.1) Sherlock Holmes exists.
(1.2) Ebenezer Scrooge exists.
(1.3) Ebenezer Scrooge is a detective.
(1.4) Sherlock Holmes is a detective.
(1.5) Ebenezer Scrooge does not exist.
(1.6) According to the novel, *A Christmas Carol*, Ebenezer Scrooge exists.
(1.7) According to the Conan Doyle stories, Sherlock Holmes is a detective.

Common sense tells us that sentences (1.1)–(1.7) are meaningful—that is that they express propositions—and that (at least some of) these propositions have a classical truth value. Sentences (1.1) and (1.2), for example, express false propositions; sentences (1.5)–(1.7), on the other hand, express true propositions. It has sometimes been maintained that the direct reference theorist cannot agree (cf. Katz, *op. cit.*). The worry can be put as follows: sentences containing a (non-mentioned) occurrence of a nonreferring name obtain their meaning compositionally; their meaning is determined in some way by the meaning of their parts and the manner in which they are combined. Each part of a sentence plays some role in determining which proposition is expressed by that sentence. Thus, such a sentence expresses a proposition only if each of its parts has some meaning. But according to the direct reference theorist, the meaning or semantic value of a name just is its referent. She thus seems forced to concede that sentences containing a nonmentioned occurrence of an empty name express no proposition at all. The embarrassment for the direct reference theorist, then, is that sentences (1.1)–(1.7) do seem to express propositions.

---

[6] Nathan Salmon, "Nonexistence," *Noûs*, XXXII (1998): 277–319, here pp. 277–78.

It has often been noted, however, that the direct reference theorist can provide a simple and straightforward response to this kind of objection. She might stipulate that a simple subject/predicate sentence will be true if and only if the subject expression refers to an object (or objects) which has (have) the property expressed by the predicate; it will be false otherwise.[7] Thus, any sentence of the form '*a* is *F*' will be false if the name '*a*' has no referent. It will also be part of this view that when such sentences occur within the scope of a sentential operator, the complex sentence may be true. According to this semantic picture, sentences (1.1)–(1.4) express false propositions, for they are simple sentences with an empty subject expression. Because sentence (1.5) is simply the negation of (1.2), and because (1.2) expresses a false proposition, (1.5) will express a true proposition. Because a fictional operator is a nontruth-functional operator, it can consistently be maintained that (1.6) and (1.7) are both true, even though the embedded sentences are false. It is important to note that this suggestion is consistent with the compositionality principle; the proposition expressed by any of the above sentences is a function of the semantic values, *or lack thereof*, of the constituents of that sentence.

This view can be wedded to a theory of singular propositions. It can be maintained that declarative sentences containing a nonmentioned occurrence of a proper name express singular propositions. When a declarative sentence contains an occurrence of an empty name—a name with no semantic value—it will express a special kind of singular proposition: a gappy proposition. A gappy proposition has a propositional structure like any other singular proposition, but one of the positions is left unfilled.[8]

---

[7] Cf. T.J. Smiley, "Sense without Denotation," *Analysis*, xx (1960): 125–35.

[8] Cf. David Kaplan, "Demonstratives," in Joseph Almog, John Perry, and Howard Wettstein, eds , *Themes from Kaplan* (New York: Oxford, 1989), pp. 481–564; David Braun, "Empty Names," *Noûs*, xxvii (1993): 449–69; Fred Adams and Robert Stecker, "Vacuous Singular Terms," *Mind and Language*, ix (1994): 387–401; Graham Oppy, "The Philosophical Insignificance of Gödel's Slingshot," *Mind*, cvi (1997): 121–41, and Salmon, "Nonexistence."

Let me illustrate, using a notation that is hopefully self-explanatory. According to the gappy proposition view, (1.3) and (1.4) both generate the following gappy proposition: <GAP, being-a-detective>. Likewise, (1.1) and (1.2) generate the following proposition: <GAP, exists>. Because these propositions (like the sentences that express them) are simple, all such propositions are false. Sentence (1.5)—the negation of (1.2)—will generate the complex gappy proposition: <NOT <GAP, exists>>. Because the proposition expressed by (1.2) is false, the proposition expressed by (1.5) is true. Sentences (1.6) and (1.7) will respectively generate the following complex gappy propositions: <Fiction$_{SH}$ <GAP, being-a-detective>> and <Fiction$_{CC}$ <GAP, exists>>. It can consistently be maintained that both propositions are true. This is all as we should expect.

This popular answer to the problem outlined above has the advantage of agreeing with common sense insofar as it allows (i) that sentences containing an occurrence of an empty name are meaningful, that is, that they express a proposition, and (ii) that the proposition expressed by such sentences can have a (classical) truth value. But the view has the disadvantage of disagreeing with common sense insofar as the theory entails that any two simple sentences, differing only with respect to the empty name contained therein, will express the same proposition. As a consequence, an adherent of this view appears committed to the thesis that substitution of one empty name for another in nonextensional contexts will always preserve the truth value of those sentences. But this does not seem right. Consider again sentences (1.6) and (1.7):

(1.6) According to the novel, *A Christmas Carol*, Ebenezer Scrooge exists.
(1.7) According to the Conan Doyle stories, Sherlock Holmes is a detective.

And compare them to the following:

(1.8) According to the novel, *A Christmas Carol*, Sherlock Holmes exists.
(1.9) According to the Conan Doyle stories, Ebenezer Scrooge is a detective.

Recall that, if we accept the direct reference theorist's solution to the original problem, the embedded simple sentence in (1.6) expresses the very same proposition as that expressed by the embedded simple sentence in (1.8). Likewise, the embedded sentence in (1.7) expresses the same proposition as that expressed by the embedded sentence in (1.9). As a consequence, (1.6) and (1.8) should express the very same proposition, and therefore have the very same truth value; so too should (1.7) and (1.9). But nothing could be further from the truth. While (1.6) is true, (1.8) is false; (1.7) is true, yet (1.9) is false.[9] This, I maintain, is the main difficulty empty names present the direct reference theorist.

The problem, though, is not a problem for the direct reference theorist alone. It is a difficulty for the descriptivist also. Or more precisely: there is a similar problem for any descriptivist who embraces the plausible thesis that names are rigid designators.

II. DESCRIPTIVISM AND THE RIGIDITY THESIS

Kripke, in his groundbreaking monograph *Naming and Necessity*,[10] argued that proper names, as they occur in natural languages, are rigid designators. A term *t* is a rigid designator if and only if *t* designates the very same object in every possible world in which *t* designates anything at all. This definition of rigidity, however, is multiply ambiguous; it allows us to distinguish a number of different senses in which a designating term can be rigid. We can make them explicit as follows:

> A term *t* is *vacuously* rigid if *t* designates no object in any possible world; *t* is *nonvacuously* rigid if it designates some object *o* in some possible world, and in no world designates an object other than *o*. If *t* designates *o* in every possible world, whether or not *o* exists there, we call *t obstinately* rigid. If *t* designates *o* in all and only those worlds where *o* exists, we call *t persistently* rigid. If *t* designates *o* in only some of the worlds in which *o* exists, and in no others, we call *t tenaciously* rigid.

Which notion of rigidity does the semantic theorist have in mind when she defends the thesis that names are rigid designators? She is not claiming that names are tenaciously rigid. Proper names are not tenaciously rigid,[11] but there are other designators that are good candidates for this title. Consider the definite description 'the father of Saul Kripke'. If we embrace the plausible assumption that Kripke could not have had a different father, it might be supposed that this description is tenaciously rigid. The description 'the father of Saul Kripke' designates the same individual, Myer Kripke, in every possible world in which it designates anything at all. It does not designate him in those worlds in which he does not exist, and (perhaps) it does not designate him in those worlds in which he is not a father.

---

[9] Fictional contexts, therefore, are (or appear to be) opaque contexts. It is a well-known fact that opaque contexts present a problem for direct reference theorists. But the literature on the subject has tended to focus on propositional attitude contexts rather than fictional contexts, for it appears that in such contexts, substitution of names—that is, either empty names or referring names—with the same semantic value will not always preserve the truth of what is said. It might be thought, therefore, that empty names per se, do not present a special problem for direct reference theories; perhaps the problem presented above is merely an instance of a more general problem. Nonetheless, there are two reasons for thinking sentences containing an occurrence of an empty name uttered in such contexts are of special interest: (i) if our attention is focused on empty names rather than referring names, we notice a similar substitutivity problem arises for the descriptivist, and (ii) various responses proffered by direct reference theorists to substitutivity problems in propositional attitude contexts—cf. especially Salmon, *Frege's Puzzle* (Cambridge: MIT, 1986); and Soames, "Direct Reference, Propositional Attitudes, and Semantic Content," *Philosophical Topics*, XV (1987): 47–87—are less plausible when we consider empty names in fictional contexts.

[10] "Naming and Necessity," was first published in Gilbert Harman and Donald Davidson, eds., *The Semantics of Natural Language* (Boston: Reidel, 1972); reprinted in monograph form (with preface added) (Cambridge: Harvard, 1980). All page references are to the later edition; hereafter NN.

The semantic theorist is thus forced to choose between the thesis that proper names are obstinately rigid designators and the thesis that proper names are persistently rigid designators. The overwhelming majority of philosophers who discuss the issue suggest that the right notion of rigidity is the first: proper names are obstinately rigid designators.[12] They reason as follows. Consider the sentence 'David Kaplan might not have existed'. Since Kaplan is not a necessary existent, this sentence must be true (that is, express a true proposition). According to the accepted semantics for such statements, this sentence is true only if there is a possible world in which 'David Kaplan does not exist' is true. But if proper names were only persistently rigid—if 'David Kaplan' designated nothing in those worlds in which he did not exist—then the sentence ‘David Kaplan does not exist’ would not be true at any world. Conclusion: proper names are obstinately rigid.[13]

---

[11] Or: most names are not tenaciously rigid. David Lewis has pointed out to me that there is a case to be made for the hypothesis that some proper names are tenaciously rigid. Perhaps 'Elizabeth II' is tenaciously rigid, if that name (as used by us in this world) does not designate Elizabeth in worlds where she exists but never takes the throne. Perhaps the new names taken by bishops when they become popes, or the name 'Great Leader' as used to designate Kim Jong-Il of North Korea, are also tenaciously rigid.

[12] Cf. Kaplan, "Demonstratives," and "Bob and Carol and Ted and Alice," in Jaakko Hintikka, Julius Moravcsik, and Patrick Suppes, eds., *Approaches to Natural Language* (Boston: Reidel, 1973), pp. 490–518; Colin McGinn, "Rigid Designation and Semantic Value," *Philosophical Quarterly*, xxxii (1982): 97–115; Esa Saarinen, "How to Frege a Russell-Kaplan," *Noûs*, xvi, 2 (May 1982): 257–276; Salmon, *Reference and Essence* (Princeton: University Press, 1981); and A.D. Smith, "Rigidity and Scope," *Mind*, xciii (1984): 177–93, and "Semantical Considerations on Rigid Designation," *Mind*, xcvi (1987): 83–92. There are, however, two notable exceptions. Kripke categorically denies that the view of proper names that he intended to defend is the view that names are obstinately rigid. But he does not, as is often maintained, mean to suggest that they are persistently rigid either Kripke intended to be deliberately neutral on this issue. Kripke's views on this matter are made clear in note 21 of NN, and in Kaplan, "Demonstratives," pp. 569–71. Robert Steinman—"Kripke Rigidity versus Kaplan Rigidity," *Mind*, xciv (1985): 431–42—does defend the view that proper names are persistently rigid and not obstinately rigid A good response to Steinman is to be found in Smith, "Semantical Considerations on Rigid Designation."

It is worth noting, however, that there is a further way of characterizing the notion of rigidity to be found in the literature, a notion that is quite distinct from the three notions of rigidity defined above. It may be stated as follows:

> A term *t* is *insularly* rigid if and only if an utterance of a simple sentence *S* that contains a nonmentioned occurrence of *t* is such that *t* designates the same individual as it would if *S* appeared within the context of a modal operator.

It is important to note that insular rigidity is not equivalent to any of the three other notions of rigidity defined above. Nor is it a way of disambiguating Kripke's notion. Indeed, Kripke is explicit, in a number of places, that insular rigidity is a much weaker notion than the notion of rigidity that he had in mind. Nonetheless, many descriptivists have embraced this weaker notion, suggesting that it is consistent with the evidence for the rigidity hypothesis, and therefore any designator that meets this condition is a good enough candidate for the name ‘rigid designator’.

Kripke mounts a concerted attack against the ‘Frege-Russell’ theory of names in *Naming and Necessity*. There he proffers a variety of arguments designed to persuade us of the untenability of such theories. Perhaps the most influential of these—the modal argument—can be reformulated as follows[14]:

Let ‘*N*’ be some name and ‘the *F*’ a description that uniquely picks out *N*—a purported candidate for the meaning of ‘*N*’.

(i) ‘The *F* might not have been an *F*’ is ambiguous.
(ii) If ‘*N*’ were used to mean ‘the *F*’, then ‘*N* might not have been an *F*’ would also be ambiguous.
(iii) But ‘*N* might not have been an *F*’ is unambiguously true.
(iv) Therefore, ‘*N*’ is not used to mean ‘the *F*’.[15]

---

[13] Of course, such an argument hardly seems compelling in light of the considerations outlined in the previous section. If ‘Kaplan’ designates nothing in worlds in which he does not exist, and as a consequence ‘Kaplan exists’ expresses a gappy proposition in those worlds, then the proposition would be false in those worlds, and moreover its negation would be true.

[14] It is interesting to note that Kripke himself does not advance an argument of this form. However, many commentators on *Naming and Necessity*—influenced by Brian Loar, "The Semantics of Singular Terms," *Philosophical Studies*, XXX (1976): 353–78—have casually attributed this argument to him.

[15] The argument requires some clarification. If we are to assume that the particulars that are named can possess essential properties, the uniquely identifying description associated with the name had better not pick out some such property. But as Kripke observes, this is hardly ever the case. It is rare (if not impossible) that we know enough

The objection can be applied to those who adopt a theory in which names are abbreviations of *a set or cluster* of descriptions.[16] According to the cluster theory, it might turn out that some of the descriptions associated with a name are false, and yet the name still refers. But for reference to occur it is necessary that some critical number of associated descriptions be true—or almost true. Kripke, quite rightly, has qualms about this modified descriptivism. Consider the following:

> Most of the things commonly attributed to Aristotle are things that Aristotle might not have done at all. In a situation in which he didn't do them, we would describe that as a situation in which *Aristotle* didn't do them.... Not only is it true *of* the man Aristotle that he might not have gone into pedagogy; it is also true that we use the term 'Aristotle' in such a way that, in thinking of a counterfactual situation in which Aristotle didn't go to *any* of the fields and do *any* of the achievements we commonly attribute to him, still we would say that was a situation in which *Aristotle* did not do these things (NN 61–62).

The modal argument against descriptivism is intimately associated with Kripke's doctrine that names are rigid designators. How so? It should be noticed that ordinary definite descriptions can give rise to scope ambiguities in modal contexts. That is why the first premise of the argument is true. Names, on the other hand, do not inherit the same kind of ambiguity. That is why the third premise of the argument is true. That names do not give rise to the same kind of ambiguity is entailed by the fact that names are rigid designators.

Some philosophers, in response to Kripke's argument, maintain that while the modal argument rationally compels the semantic theorist to embrace the thesis that names are rigid designators in some sense, it in no way counts against descriptivism. The semantic evidence is consistent, it is suggested, with the thesis that names are merely insularly rigid designators. If this diagnosis of the semantic evidence in support of the rigidity thesis is right, the descriptivist has at hand a response to the objection Kripke and others have mounted against their semantic theory.

The response runs as follows. Kripke is right to maintain that names are rigid designators; but they are only insularly rigid. This is because names are equivalent to descriptions that are semantically required to take wide scope over modal operators occurring in the same sentence. Thus, Kripke's objection does not get a hold. If there really is a

---

about the essential properties of any given individual that our related description will refer uniquely to that individual.

[16] Searle, "Proper Names," *Mind*, LXVII (1958): 166–73.

convention in natural language that proper names, unlike definite descriptions, always take wide scope over modal operators that occur in the same sentence, then the second premise of the modal argument is false. Call this view *wide-scope descriptivism*.[17]

It seems that this is an adequate response to the modal argument, as it is stated above. Nonetheless, it has been maintained by Kripke and others[18] that the argument can be modified in such a way as to make scope distinctions irrelevant. The modified version of the argument runs as follows:

Let '*N*' be some name and 'the *F*' a description that uniquely picks out *N*—a purported candidate for the meaning of '*N*'.

(i) 'The *F* is an *F*' is true in every world containing a unique *F*.
(ii) If '*N*' were used to mean 'the *F*', then '*N* is an *F*' would also be true in every world containing a unique *F*.
(iii) But '*N* is an *F*' is not true in every world containing a unique *F*.
(iv) Therefore, '*N*' is not used to mean 'the *F*'.

This version of the modal argument has a (supposed) advantage over the initial version of the argument insofar as the sentences we are asked to evaluate are simple sentences, and therefore cannot give rise to the scope ambiguities on which the initial version of the argument traded. Thus, the response to the initial version of the modal argument proffered by the wide-scope descriptivist cannot be applied to the modified version of the argument. The reason, it is suggested, is that names, unlike descriptions, are obstinately rigid, and not merely insularly rigid. Thus, descriptivism is false.

Whether or not this modified version of the modal argument decisively refutes wide-scope descriptivism is a matter of some controversy. Nonetheless, I do not care to adjudicate between the two sides here.[19]

---

[17] This response to Kripke has been proposed or embraced by an overwhelming number of philosophers sympathetic to the descriptivist position—see, for example, B. Brody, "Kripke on Proper Names," *Midwest Studies in Philosophy*, II (1977): 75–80, Michael Dummett, *Frege: Philosophy of Language* (Cambridge: Harvard, 1973); Brian Garrett, "More on Rigidity and Scope," *Logique et Analyse* (1984): 97–102, Loar; David Sosa, "Rigidity in the Scope of Russell's Theory," *Noûs*, XXXV (2001): 1–38; and Paul Yu, "The Modal Argument against Description Theories of Names," *Analysis*, XL (1980): 208–09. For criticism of this view, see NN 11–12; and Soames, "The Modal Argument: Wide Scope and Rigidified Descriptions," *Noûs*, XXXII (1998): 1–22.

[18] See the preface to NN, and James Hudson and Michael Tye, "Proper Names and Definite Descriptions with Widest Possible Scope," *Analysis*, XL (1980): 63–64.

[19] Some philosophers have responded to arguments like this by attempting to show that the premises of the argument are equivalent to sentences which do give rise to scope ambiguities, and so therefore the argument is unsound for exactly the same reason that the initial argument is (cf. Graeme Forbes, *Languages of Possibility* (New York: Blackwell, 1989); Sosa; and Yu).

Instead, I merely want to point out that even if this argument were successful against the wide-scope descriptivist, it would not decisively refute other varieties of descriptivism. Against Kripke, a descriptivist can still maintain that the relevant descriptions can be rigidified in the appropriate way so as to avoid any commitment to the conclusion of this new version of the modal argument. How might the descriptivist do this? Let me review one such suggestion.

The descriptivist I have in mind will suggest that the descriptions offered by Russell, Frege, and others be modified in such a way as to index them to a possible world. Perhaps the most straightforward way to do this is to incorporate an actuality operator into the relevant descriptions. Thus, it might be said that 'Aristotle', for example, means something like the *actual* pupil of Plato and the *actual* teacher of Alexander the Great. This new kind of description is a kind of obstinately rigid designator. So where does Kripke's modal argument go wrong? If the definite description used in either version of the argument is a world-indexed description, premise (i) will be false. Call this view *world-indexed descriptivism*.[20]

We are now in a position to demonstrate that if the descriptivist seeks to accommodate Kripke's modal intuitions—and thereby respect the evidence for the rigidity of proper names—she will, somewhere along the line, find herself in trouble. Where we locate the trouble will depend on which variety of descriptivism—world-indexed descriptivism or wide-scope descriptivism—is under consideration.

III. WORLD-INDEXED DESCRIPTIVISM

Consider the following three theses about Santa Claus:

(3.1) Santa Claus does not exist.
(3.2) Santa Claus might have existed.
(3.3) It is possible that Santa Claus has a beard, lives at the north pole, climbs down chimneys to give children presents on Christmas morning, and so forth.

As is well known, the descriptivist (like the direct reference theorist) can account for our transparent intuition that (3.1) is true. She has at hand something like the following plausible paraphrase of (3.1) where '$F!$' is a predicate like 'is a *unique* big fat jolly man who has a beard and lives at the north pole and so forth'; 'A' is the modal operator 'actually'[21]:

---

[20] This response to Kripke is embraced by philosophers such as Leonard Linsky, *Names and Descriptions* (Chicago University Press, 1977); Plantinga; and Frank Jackson, "Reference and Description Revisited," *Philosophical Perspectives*, XII (1998): 201–18. A thorough and sympathetic presentation of the idea is given in Jason Stanley, "Names and Rigid Designation," in Bob Hale and Crispin Wright, ed., *A Companion to the Philosophy of Language* (Cambridge: Blackwell, 1997), pp. 555–85. For criticism of this view, see Salmon, *Reference and Essence*, pp. 26ff. and Soames, "The Modal Argument."

(3.1A) $\sim\exists x\,(\mathrm{A}F!x)$

Moreover, it is clear that the descriptivist can paraphrase a vast array of other sentences containing empty names in an intuitively plausible way. This advantage, though, has all too often been overemphasized. Let us see why.

Consider first world-indexed descriptivism. The adherent of this view will paraphrase (3.2) as something like,

(3.2A) $\Diamond\exists x\,(\mathrm{A}F!x)$

On this account, (3.2) is false. If it is false that *in the actual world,* there is an x that has a beard, lives at the north pole,..., then it is also false that there is a possible world such that: *in the actual world* there is an x that has a beard, lives at the north pole,.... Santa Claus, then, could not have existed according to world-indexed descriptivism.

It seems to me that this is a prima facie cause for concern. It is a commonplace that Santa Claus might have existed.[22] The modality here is not an epistemic one. We continue to be disposed to assert (3.2) once we learn that Santa does not in fact exist. There may of course be some philosophers, deeply entrenched in their own theoretical views, who have forgotten that this is the common-sense view. But it should be clear from Kripke's own comments on the subject that to deny (3.2) does some violence to our intuitions. Consider the following:

---

[21] The open formula $F!x$ is simply shorthand for $(Fx \,\&\, \forall y\,(Fy \rightarrow x = y))$ and can be read as 'x is a unique F'. See J. Crossley and I.L. Humberstone, "The Logic of 'Actually'," *Reports on Mathematical Logic*, VIII (1977): 11–29, for a detailed explanation of the logical behavior of the actuality operator: 'A'. Of course the wide-scope descriptivist's translation of (3.1) will omit all occurrences of 'A' (cf. paraphrase (3.1B)).

[22] In response to the argument I present below against world-indexed descriptivism, it has sometimes been suggested to me that (3.2) should be taken to mean something like:

(3.2*) 'Santa Claus' might have referred.

If it is, the trouble disappears. But the suggestion seems implausible to me, for on this view, negative existentials that "use" different names cannot have the same meaning—they automatically express different propositions. Intuitively, though, it seems as though the sentence 'Lè Pere Noël n'existe pas' can be used to express exactly the same proposition as that expressed by an utterance of (3.1).

> Holmes does not exist, but in other states of affairs, he would have existed.[23]

And later, after a change of opinion:

> Some of the views I have are views which may at first glance strike some as *obviously wrong*. My favorite example is this (which I probably won't defend in these lectures—for one thing *it doesn't ever convince anyone*): It is a common claim in contemporary philosophy that there are certain predicates (and names) which, though they are in fact empty—have null extension—have it as a matter of contingent fact.... [A]n example often given is the example of a unicorn. So it is said that though we have all found that there are no unicorns, of course there might have been unicorns. Under certain circumstances there would have been unicorns. And this is an example of something I think is not the case... Now I don't know if I'm going to have time to defend this particular view, *but it's an example of a surprising one*.... So, *some of my opinions are somewhat surprising* (NN 23–24, my emphasis).

But commonplaces can be challenged. We often find good reasons to reject common-sense intuitions, reasons that are independent of our current theoretical concerns. It is ironic that Kripke himself has provided the adherent of world-indexed descriptivism with what might seem to be the required justification.[24]

Kripke puts forward a number of important considerations that are meant to persuade us that sentences like (3.2) are in fact false. Kripke's discussion focuses for the most part on sentences that contain a nonreferring natural kind term (like 'unicorn'). But the most salient of these considerations pertain to sentences that contain an instance of a nonreferring proper name. Kripke suggests that no matter what cluster of properties we have in mind when we use terms like 'unicorn' or 'Santa Claus', if the term does not refer to any actual object it cannot refer to any possible object either. Why? Because it is much harder—if not impossible—to find a description that uniquely picks out a merely possible individual (or natural kind) than it is to find one that picks out an actual individual (or natural kind). Any number of nonactual characters could fit the description found in the Santa Claus story. Indeed, there may be a single world in which a plurality of individuals are such that they live at the north pole, have a beard, own flying reindeer, give out presents early on Christmas morning, and so forth. But recall that the descriptivist and direct reference theorist alike are in agreement that names are rigid designators. Thus names can refer to only one (possible) individual. Because there is no fact of the matter to appeal to when deciding which of these hypothetical characters would have been Santa Claus, it seems wrong to suggest that Santa might have existed at all.

---

[23] Kripke, "Semantical Considerations on Modal Logic," *Acta Philosophica Fennica*, xvi (1963): 83–94, here p. 85.

[24] Kaplan, in "Bob and Carroll and Ted and Alice," makes essentially the same point in Appendix XI, pp. 505–08.

Kripke's argument might seem compelling, and the adherent of world-indexed descriptivism may well be happy to take it on board. No matter what one thinks of the argument, though, it is not meant to show that it is obvious that (3.2) is false. Indeed, Kripke himself says that his "opinions are somewhat surprising." Thus, there remains a residual worry for the descriptivist, a worry that does not pertain to the direct reference theorist.

One of the main tenets of descriptivism, it will be recalled, is that names have senses and whenever we are in a position to use a name we must have the sense associated with it in mind. As has already been noted, the sense of a name like 'Santa Claus', according to world-indexed descriptivism, can be captured by something like the following description: 'The *actual* jolly fat man who lives at the north pole, has a beard, and so forth'. The worry we should have, then, is that once we come to realize that there is no individual who fits that description, and therefore that Santa Claus does not in fact exist, we should be in a position to infer immediately that Santa could not possibly have existed. To fail to make this observation—if we really had this description in mind whenever we used the name 'Santa Claus'—would be to suffer from some rational failing or other, perhaps because we do not understand the seemingly obvious logical relation that stands between sentences (3.1) and (3.2), perhaps because we do not pay attention to (or understand) our own use of the modal operator 'actually'. But most people are disposed to assert both (3.1) and (3.2), and it seems to me that we should be suspicious of any theory that attributes such widespread stupidity.[25]

---

[25] An anonymous referee has pointed out that this objection to the wide-scope descriptivist might not be as costly as I make it appear. It is commonly accepted, for example, that no human is logically omniscient; that is, humans do not know all of the logical consequences of what they know. But this is not a reason for attributing widespread stupidity; to be ignorant of the logical entailments of one's beliefs is not a way to be dim-witted or irrational. It might be thought, therefore, that the wide-scope descriptivist can take succor from this observation. Perhaps our failure to notice the necessity of Santa Claus's nonexistence after recognizing that there is no Santa is simply an instance of this general phenomenon. It seems to me, however, that my objection is relevantly different in two respects. First, if the world-indexed descriptivist is right about the content of our beliefs, most people who know that Santa does not exist not only fail to believe a logical consequence of that belief, they also believe the negation of such a consequence. Second, if the wide-scope descripti-

IV. WIDE-SCOPE DESCRIPTIVISM

It might therefore be hoped that trouble can be avoided if we adopt instead a kind of wide-scope descriptivism. An adherent of this kind of descriptivism will be in agreement with common sense insofar as she is committed to the truth of theses (3.1) and (3.2). She will presumably paraphrase each as follows:

(3.1B) $\sim\exists x\,(F!x)$
(3.2B) $\Diamond\exists x\,(F!x)$[28]

There is, however, a problem for an adherent of this variety of descriptivism, a problem analogous to that mounted against the world-indexed descriptivist.

If it is possible that Santa Claus exists, then it should also be possible that Santa Claus is *an F*, where '*the F*' is the description capturing the meaning of the name 'Santa Claus'. If the second thesis about Santa is true, the third should be true also. For it is analytic that it is possible that there exists an *F* only if it is possible that there exists an *F* that is *F*.

The wide-scope descriptivist, though, cannot agree. Recall that the wide-scope descriptivist seeks to endorse the rigidity thesis by maintaining that the logical form of sentences containing a proper name and a modal operator are such that the description (of which the name is an abbreviation) invariably takes wide scope. Thus, her translation of (3.3) should be as follows:

(3.3B) $\exists x\,(F!x\ \&\ \Diamond Fx)$

But (3.3B) is clearly false. We have learned that there is no individual that is picked out by the Santa Claus description, yet (3.3B) says there is. The account must therefore be rejected.

The arguments outlined in the last two sections spell trouble for the descriptivist. But the arguments are not knock-down arguments, and the keen reader will no doubt anticipate a number of potential escape routes available to the descriptivist. In the sequel I will consider two such rejoinders, one inspired by Russell, the other by Frege. My

---

vist were right, the consequence should be a trivial one. And it is a minimal requirement of rational agents that they recognize the trivial consequences of their beliefs, even if they are not required to recognize every logical consequence

[28] Notice that there is no room for any scope ambiguities here, for 'existence' is not treated as a predicate. If, however, it were, sentence (3.2) would instead be paraphrased as follows: $\exists x\,(F!x\ \&\ \Diamond Ex)$. Because there is in fact no *F*, an adherent of this view is thereby committed to the view that (3.2) is false. Thus, she is vulnerable to exactly the same objection mounted against the world-indexed descriptivist in section III.

task here is not evaluative; my aim is merely to point out that however the descriptivist might respond, a parallel and equally plausible response is open to the direct reference theorist.

V. THE HYBRID VIEW

If what I have said thus far is correct, there is trouble in the offing for the descriptivist who is committed to the thesis that all names—both those that refer and those that do not—are rigid designators. But perhaps she need not commit herself to such a thesis; perhaps all that the preceding discussion shows is that empty names cannot be rigid designators even if ordinary referring names are.

The adherent of a hybrid variety of world-indexed descriptivism will suggest that a referring name will be captured by a rigidified description of the form 'the *actual F*', whereas a nonreferring name will be captured by a nonrigidified description of the form 'the *F*'. Those who embrace a hybrid variety of wide-scope descriptivism, on the other hand, will suggest that the logical form of any sentence containing a proper name and a modal operator will vary depending on whether or not the name refers. If it does, the description associated with the name will take wide scope; if it does not, the description will take narrow scope.

The semantic theorist who accepts this kind of hypothesis is in a position to reconcile her theory with some of our firmest intuitions about proper names. She will agree with all of Kripke's modal intuitions, except when it comes to the case of empty names. When we restrict our attention to the case of empty names only, premise (3) of Kripke's modal argument—the premise that any sentence of the form '*N* might not have been the *F*' is unambiguously true—loses much of its plausibility. For when we come to consider ordinary and mundane varieties of empty names it is not unreasonable to suppose that all (or most) of the descriptions we associate with these names are, so to speak, essential to the characters in question.

Moreover, the proponent of this style of account will be able to accommodate our intuitions that Santa Claus does not exist, that it is possible that he exists, and that if he had existed, he might have been a kind fat jolly man with a beard. An adherent of either variety of descriptivism about referring names—wide-scope descriptivism or world-indexed descriptivism—who also advocates the hybrid view about proper names in general, will suggest that theses (3.1)–(3.3) have the following logical form:

(3.1B) $\sim\exists x\,(F|x)$
(3.2B) $\Diamond\exists x\,(F|x)$
(3.3H) $\Diamond\exists x\,(F|x\ \&\ Fx)$

On this account, then, theses (3.1)–(3.3) are all true. Indeed, thesis (3.3) is analytic. This is all as we should expect.

Such a proposal commits us to the view that the difference between the semantic value of an empty name and that of a referring name is a difference in kind. This difference is postulated because of the problems outlined in section III and section IV. Having made such a concession to the descriptivist, however, we must be prepared to make a similar allowance for the direct reference theorist.

A number of philosophers sympathetic to the theory of direct reference, and bothered by the difficulties empty names present this theory, have suggested that so-called empty names are not really names at all.[27] Instead, such terms are treated as disguised descriptions. Because they are not singular terms, it is maintained, the Millian is not responsible for interpreting sentences containing an occurrence of an empty 'name'.[28]

Whether we choose to call terms like 'Santa Claus', 'Vulcan', and 'Sherlock Holmes' names or not seems of little relevance. The important point to note in the present context is that, on such a view, the semantic value of these terms is of a significantly different kind to that of referring names. Thus, Currie and Ryckman both offer the direct reference variant of the hybrid view. To illustrate, consider the following matrix, representing the alternative views.

	Referring Names	*Empty Names*
Descriptivism	Rigidified Description	Flaccid Description
Direct Reference	An Individual	Flaccid Description

Whatever the merits and disadvantages of these views, their respective treatment of empty names is exactly the same in both cases; both stand or fall together on this issue.[29]

---

[27] See, for example, Ryckman; and Gregory Currie, *The Nature of Fiction* (New York: Cambridge, 1990), chapter 4.

[28] Russell's official view about proper names was very close to the theory propounded by the hybrid variety of direct reference theory. For Russell, though, the dichotomy is slightly different; instead of referring names, Russell considers the class of names we are absolutely certain refer. These names, suggests Russell, are directly referential. All others are abbreviated definite descriptions.

[29] Lewis has suggested to me that there is an alternative kind of hybrid view that might be proposed by the descriptivist which would evade the problems outlined in section III and section IV. Lewis suggests that descriptions generally, and proper names in particular, are once and for all ambiguous between a rigidified and unrigidified sense. We therefore have to rely on facts about the conversational context to determine which kind of proposition the speaker intends to express. This kind of view is significantly different than the position I outline above insofar as the adherent of this kind of view has an easy response to Kripke's modal argument (outlined in section II). Unlike the descriptivist of any variety considered thus far, she will deny premise (3) of the modal argument—the premise that sentences like 'Aristotle might

VI. THE REALIST VIEW

Recall that a descriptivist says that names are just abbreviated definite descriptions and so a sentence containing a proper name has the same logical structure as a sentence containing a definite description. We may, however, distinguish two main approaches to the problem of how best to analyze the logical structure of sentences containing a definite description.

The first, Fregean, approach treats definite descriptions as singular terms, so that '*a* is *G*' and 'the *F* is *G*' genuinely have the same sentential structure, though the latter involves a complex singular term. One virtue of this account is that the logical form of a sentence like 'the cat in the hat came to play' reflects its surface grammatical form.

The second, Russellian, approach denies that definite descriptions are singular terms. Instead, definite descriptions are treated as a species of singular restricted quantifier, binding variables rather than substituting for them, so that '*a* is *G*' and 'the *F* is *G*' differ in logical form. More precisely, a definite description of the form 'the *F* is *G*' is paraphrased as follows:

(6.1) $\exists x\ (F!x\ \&\ Gx)$

When outlining the difficulties empty names present the descriptivist, I have assumed (with Russell) that descriptions are a species of singular restricted quantifier rather than singular terms. This is not because I believe the objections do not have force against the Fregean view. The reason for preferring Russell's approach over Frege's starts from the observation that because the Fregean approach treats definite descriptions as singular terms, it seems to face a similar problem to that outlined in section I against the direct reference theorist. Like the direct reference theorist, Frege is not committed to the view that sentences containing an occurrence of an (apparently) empty singular term have no meaning. But Frege, at first blush, seems to be committed to the view that all such sentences have no (classical) truth value. For according to Frege, a declarative sentence has a (classical) truth value only if all singular terms contained therein refer.

---

not have been a philosopher' are unambiguously true. Moreover, a descriptivist of this variety will not have any difficulty accounting for our intuitions about sentences containing an occurrence of an empty name. It is important to note, however, that if one finds this kind of view plausible, a similar response to the problem of empty names is open to the direct reference theorist. She can likewise suggest that proper names are once and for all ambiguous (between a descriptivist and directly referential sense).

Frege maintained that the problem could be solved in a variety of different ways. But in order to preserve our intuition that all such sentences express a proposition with a classical truth value, Frege suggested that a singular term "shall in fact designate an object, and that no new sign shall be introduced as a proper name without being secured a reference" (*op. cit.*, p. 70). The way Frege proposes to do this is to insist that an arbitrary object be provided if necessary. He says that a singular term "must always be assured of reference by means of a special stipulation, for example by the convention that 0 shall count as its reference when the concept applies to no object or more than one" (*op. cit.*, p. 70).

Whatever the merits of this arbitrarily-chosen-object view, it is inadequate as an account of the semantics of (so-called) empty names in natural language.[30] There are at least two problems with the approach: (i) the view appears to give the wrong distribution of truth values to the following kinds of sentences: 'Sherlock Holmes is a number', 'The name "Bart Simpson" has a referent', 'Johnny believes that Santa exists because he believes that the number 0 exists', all of which are (or appear to be) true according to this hypothesis, and, perhaps more importantly, (ii) an adherent of this approach appears committed to the view that substituting one "empty" term for another in any sentence *S* can never change the truth value of the proposition expressed by *S*, for the referent in each case will be the same.[31]

One way out of this difficulty is to suggest first that there is an abundance of such objects and second, that the individuals to which these so-called empty names refer is no arbitrary matter. Perhaps such names pick out various Meinongian nonexistent objects, or perhaps they denote abstract particulars, or perhaps they refer to some other kind of thing altogether.[illegible] But no matter how the realist's thesis is spelled out, we have a solution to the problem at hand. Santa Claus might have existed (because there is such an entity); and if he had, he would have been a kind fat jolly man (or at least, he would have been so represented). Moreover, we can leave room for a sense in which claims like 'There is no Santa Claus' are true, so long as we understand such quantification as implicitly restricted.

---

[30] Frege, of course, never maintained that this was a way of explaining our intuitions about empty names as they occur in natural languages. It was a technical solution to a technical difficulty, that is, a suggestion for side-stepping this problem when we come to construct an artificial language. Nonetheless, other philosophers, influenced by the work of Frege (for example, Rudolf Carnap, Alonzo Church, Michael Dummett, and W.V. Quine), have since arrogated to themselves the proposed solution, and put it forward as a way of understanding natural languages.

[31] Frege himself solved the problem, at least in part, by suggesting that in some nonextensional contexts names and descriptions change their mode of reference. Instead of referring to an individual, such terms come to refer to their sense. I do not consider the merits of such a proposal here, however, because (i) it does not explain why sentences like 'Sherlock Holmes is a number' have the truth value they do; and (ii) this proposal, by itself, is of no help dodging the problems outlined in section III and section IV. For an interesting discussion about how such a suggestion might help the direct reference theorist, though, see Oppy.

Embracing this kind of realism about these supposedly nonexistent entities also brings into line the Fregean theory with the natural view that sometimes substituting one nonreferring term for another in a given sentence can change the truth value of that sentence. But recall that this was the very same problem facing the direct reference theorist. Consider again sentences (1.7) and (1.9):

(1.7) According to the Conan Doyle stories, Sherlock Holmes is a detective.

(1.9) According to the Conan Doyle stories, Ebenezer Scrooge is a detective.

If the names 'Ebenezer Scrooge' and 'Sherlock Holmes' are genuinely nonreferring, then they have the same semantic value—none. And if two names have the same semantic value, substituting one name for the other in any given sentence could not change the truth value of the sentence. As a consequence, (1.7) and (1.9) should express the very same proposition, and therefore have the very same truth value. But this is clearly not the case. While (1.7) is true, (1.9) is clearly false.

One way out of the difficulty is to deny the assumption that gives rise to the problem, the assumption that 'Sherlock Holmes' and 'Ebenezer Scrooge' have the same semantic value. The realist does just that by suggesting that such names are not denotationless terms, but names that refer to two very different abstract or Meinongian individuals.

[illegible] For independent reasons to adopt this option, see Robert Howell, "Fictional Objects: How They Are and How They Aren't," *Poetics*, VIII (1979): 129–77; Kripke, "Reference and Existence," Unpublished Manuscript Presented as The John Locke Lectures (1973); Peter Lamarque, "Fiction and Reality," in Lamarque, ed., *Philosophy and Fiction: Essays in Literary Aesthetics* (Aberdeen: University Press, 1983), pp. 52–72; Jerrold Levinson, "Making Believe," *Dialogue*, XXXII (1993): 359–74; Terence Parsons, *Nonexistent Objects* (New Haven: Yale, 1980); Peter van Inwagen, "Creatures of Fiction," *American Philosophical Quarterly*, XXIV (1977): 299–08; and Nicholas Wolterstorff, "Characters and their Names," *Poetics*, VIII (1979): 101–27. Such authors tend to focus on the case of *fictional* names. But it seems to me that the considerations used by such authors to support the thesis that such names refer apply equally well, for example, to the case of names that were introduced by mistake, when someone honestly thought they referred (for example, the name 'Vulcan' which was introduced by Le Verrier to designate the unobserved planet which caused the deviations in Mercury's orbit predicted by Newton's laws of motion and gravity).

VII. CONCLUSION

And so I rest my case, the case in favor of the hypothesis that empty names do not present a special problem for the direct reference theorist; to the extent that the direct reference theorist has a problem with empty names, the descriptivist has one also. The arguments presented, though, are not decisive. Perhaps there is a natural way to dodge the difficulty available only to the descriptivist. The fact that I have not noticed it may reflect nothing more than a lack of ingenuity on my part. But, at the very least, it should be clear that this all too common presupposition can no longer be taken for granted.

STUART BROCK

Victoria University of Wellington

DIALECTICAL DELICACIES IN THE DEBATE ABOUT FREEDOM AND ALTERNATIVE POSSIBILITIES*

It is widely assumed that the freedom necessary for moral responsibility requires alternative possibilities.[1] This intuitively compelling idea, that responsibility presupposes that we have open pathways branching into the future, is encapsulated in the *principle of alternative possibilities* (PAP):

> PAP: A person is morally responsible for what she has done only if she could have done otherwise.[2]

There are, however, various ways to challenge PAP.[3] One invokes famed "Frankfurt-type" examples, so named after Harry Frankfurt's presentation of them. Such examples contain a fail-safe mechanism that does not play any role in the relevant agent's actual deliberations, choices, and behavior but that ensures that the agent deliberates, chooses, and behaves in just the way in which he does. In his original examples, Frankfurt relied upon a counterfactual intervener who monitored the agent and intervened in the agent's behavior should the agent show any tendency to do other than what the intervener desires the agent to do.[4] In subsequent refinements, if intervention is called for, the intervener relies on a "prior-sign" as a cue for such intervention. To illustrate, consider this *prior-sign Frankfurt example*:

---

* We would like to thank John Fischer and Al Mele for their excellent comments and advice

[1] Philosophers as disparate as Hume and Kant shared this view. For just a few of the many contemporary examples, see Carl Ginet, *On Action* (New York: Cambridge, 1990); Robert Kane, *The Significance of Free Will* (New York: Oxford, 1996); David Lewis, "Are We Free to Break the Laws?" *Theoria*, XLVII (1981): 113–21; and Peter van Inwagen, *An Essay on Free Will* (New York: Oxford, 1983). As we shall make clear, some argue that moral responsibility requires freedom, but not freedom involving alternative possibilities. A minority position challenges the very basic assumption that moral responsibility requires freedom. For example, in *Moral Responsibility and Persons* (Philadelphia: Temple, 1992), Eugene Schlossberger rejects the view that responsibility is a function of the agent's acting freely—see, for example, pp. 4–7, 117–18. Similar ideas are suggested in Nomy Arpaly's *Unprincipled Virtue* (New York. Oxford, 2003), pp. 149–79.

[2] See Harry G. Frankfurt, "Alternate Possibilities and Moral Responsibility," this JOURNAL, LXVI, 23 (December 4, 1969): 829–39, on p. 829.

[3] We shall focus upon Frankfurt's well-known approach. For an alternative, see, for instance, Michael J. Zimmerman, "Taking Luck Seriously," this JOURNAL, XCIX, 11 (November 2002): 553–76. Zimmerman argues that considerations of luck impugn PAP.

[4] See Frankfurt, pp. 835–36.

0022-362X/04/0106/299–314

> Mork believes that in his circumstances it is wrong for him to lie to prevent embarrassment. In spite of this, he decides to lie and lies anyway. As he has no excuse or justification for lying, it seems he is deserving of blame for lying. Unbeknownst to Mork, though, he could not have refrained from lying due to the presence of Mindy who has the power to read and control Mork's mind. Mindy wields this power partly in virtue of possessing the following knowledge. Had Mork been about to refrain from lying, he would have displayed some involuntary sign—a neurological pattern, N^*, in his brain—whereas if he had been about to lie, he would have displayed a different neurological pattern, N. Had Mindy detected N^*, she would have interceded in Mork's deliberations via direct stimulation of Mork's brain and would, in this way, have caused Mork to lie. But Mindy detects N, the reliable sign for Mindy that she need not show her hand at all. As Mork lies on his own, in the absence of Mindy's intervention, it seems highly reasonable that Mork acts freely and is morally blameworthy for lying despite not having alternative possibilities.

In light of such examples, Frankfurt and others have argued that alternative possibilities are *not* required for blameworthiness or responsibility in general. Call those who defend Frankfurt's argument *Frankfurt Defenders*.

A great deal turns upon Frankfurt's thought experiment.[5] If it is sound, one compelling argument for incompatibilism is defeated. This incompatibilist argument, the *Classical Incompatibilist Argument*, maintains that if determinism is true, no one can do otherwise; free will and moral responsibility require the ability to do otherwise; therefore, if determinism is true, no one has free will, and no one is morally responsible. Traditionally, compatibilists had wrestled with the Classical Incompatibilist Argument by challenging the premise that determinism rules out the ability to do otherwise. Frankfurt's approach offered compatibilists a way of conceding this premise while undermining this incompatibilist argument on other grounds. If, as Frankfurt argues, PAP is false, free will and moral responsibility do not require the ability to do otherwise.

Prior-sign examples have played such a crucial role in this debate because it has seemed to many that without prior-signs, Frankfurt's argument is transparently unconvincing. To illustrate, consider a

---

[5] Some terminology will prove helpful. Determinism is the thesis that, holding fixed the nonrelational facts of the past and the laws of nature, there is at any instant exactly one physically possible universe. Compatibilists believe that determinism is compatible with free will and moral responsibility. Incompatibilists hold that free will and moral responsibility are incompatible with determinism. Libertarians are incompatibilists who believe that determinism is false and that some human beings sometimes act of their own free will.

Frankfurt example that does not involve a prior-sign. Suppose that in the above example, Mindy is only able to intervene in Mork's conduct *after* Mork has decided (chosen, formed the intention) not to lie. Only in this way can Mindy ensure that Mork will lie (or, at least, utter a falsehood). If this is how the example goes, the advocate of PAP can simply point to the fact that when Mork freely lies on his own, even in Mindy's presence, he did have available a significant alternative possibility[6]; he could have decided (chosen, formed an intention) not to lie. So, the defender of PAP can argue, Frankfurt examples that do not involve prior-signs are not counterexamples to PAP. An effective counterexample must rule out any morally significant alternative within the voluntary control of the agent. Prior-sign examples, ex hypothesi, make it so that the ensuring conditions (that guarantee only one course of action) rule out alternatives prior to the initial moments of freely willed actions.[7]

I. THE DILEMMA DEFENSE AND A NEGLECTED HORN

One compelling objection to prior-sign examples is couched in terms of a dilemma.[8] If sign *N* is reliable, in the sense of being infallible, it can only be because states of the agent prior to the occurrence of the putatively free action are causally sufficient for this action (and the sign indicates this). But if that is the case, then a deterministic relation obtains between the prior-sign and subsequent behavior, and this begs the question against incompatibilists who believe that determinism is incompatible with freedom or responsibility. If, on the other hand, sign *N* is not infallible and is only reliable in some weaker sense, then an agent who acts freely in a Frankfurt example retains the ability to do otherwise when she acts on her own. The presence (or absence) of the prior-sign is consistent with the agent acting in a manner other than the manner in which she does. Call this defense of PAP the *Dilemma Defense*, and call those who advance it, *Dilemma Defenders*.[9]

---

[6] By a "morally significant alternative possibility" we mean to restrict our attention only to what some have called "robust" alternatives. Robust alternatives are alternatives that are within the scope of the agent's control and that can be used as a (partial) basis for explaining why an agent is morally responsible for her action (or in their absence, why an agent is not morally responsible for her action).

[7] This moment might be called the *locus* of a freely willed action, the baptismal point of human freedom or control.

[8] See Carl Ginet, "In Defense of the Principle of Alternative Possibilities: Why I Don't Find Frankfurt's Argument Convincing," *Philosophical Perspectives*, x (1996): 403–17; Kane, pp. 142–44; and David Widerker, "Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities," *Philosophical Review*, CIV (1995): 247–61.

[9] Note that the Dilemma Defense seems only to be available to incompatibilists,

The Dilemma Defense has instigated an impressive literature. Most Frankfurt Defenders have attempted to respond only to the second horn of the dilemma. They have attempted to produce examples that both rule out all morally significant alternative possibilities and that do not assume that the agents in the examples are determined.[10] With one notable exception,[11] Frankfurt Defenders have ignored the first horn of the dilemma.

In this paper, we focus principally on this horn—the horn that objects that the prior-sign cannot presuppose a deterministic relation. Our primary aim is to bring into relief at least two separate dialectical contexts that naturally arise in thinking about its presumed significance. These contexts can be sorted out by fixing upon two hitherto unnoticed points. One concerns distinct interpretations of the first horn. Another concerns the different audiences at which Frankfurt's argument can be appropriately directed. We propose that one interpretation of the first horn, the *broad interpretation*, is most reasonably suited for a dialectic in which Frankfurt's argument is directed at a philosophically *committed audience*. We also propose that a very different interpretation of the first horn, the *narrow interpretation*, is most reasonably suited for a dialectic in which Frankfurt's argument is directed at an *undecided audience*.

Distinguishing between these dialectical dynamics is significant be-

---

since it is only they who are entitled to the objection that, given one horn of the proposed dilemma, Frankfurt examples might infelicitously presume a deterministic relation.

[10] See Ishtiyaque Haji, *Moral Appraisability* (New York: Oxford, 1998), pp. 30–41; David Hunt, "Moral Responsibility and Unavoidability," *Philosophical Studies*, xcvii (2000): 195–227, and "Freedom, Foreknowledge, and Frankfurt," in Widerker and Michael McKenna, eds., *Moral Responsibility and Alternative Possibilities* (Aldershot, UK: Ashgate, 2003), pp 159–83; Alfred Mele and David Robb, "Rescuing Frankfurt-Style Cases," *Philosophical Review*, cvii (1998): 97–11, and also "BBs, Magnets, and Seesaws. The Metaphysics of Frankfurt-Style Cases," in Widerker and McKenna, eds., pp. 127–37; McKenna, "Robustness, Control, and the Demand for Morally Significant Alternatives," in Widerker and McKenna, eds., pp. 201–17; Derk Pereboom, "Alternative Possibilities and Causal Histories," *Philosophical Perspectives*, xiv (2000): 119–37, *Living without Free Will* (New York: Cambridge, 2001), and "Source Incompatibilism and Alternative Possibilities," in Widerker and McKenna, eds., pp. 185–99; and Eleonore Stump, "Libertarian Freedom and the Principle of Alternative Possibilities," in Daniel Howard-Snyder and Jeff Jordan, eds., *Faith, Freedom, and Rationality: Essays in the Philosophy of Religion* (Lanham, MD: Rowman and Littlefield, 1996), pp. 73–88, and "Moral Responsibility without Alternative Possibilities," in Widerker and McKenna, eds., pp. 139–58.

[11] See John Martin Fischer, "Recent Work on Moral Responsibility," *Ethics*, cx (1999): 93–139, "Frankfurt-Type Examples and Semi-Compatibilism," in Kane, ed., *The Oxford Handbook of Free Will* (New York: Oxford, 2002), pp. 281–308, and "Free Will and Moral Responsibility," in David Copp, ed., *Oxford Handbook of Ethical Theory* (New York: Oxford, forthcoming).

cause a response to the first horn depends crucially upon which dynamic is presupposed. A response seemingly effective on one interpretation misses the mark on the other. Furthermore, uncovering these different dialectical contexts will help to unearth and assess what Dilemma Defenders are centrally committed to. We propose that, on one interpretation, the Dilemma Defender is obliged to entertain the possibility that determinism rules out free will for reasons that have nothing to do with alternative possibilities. We propose that, on a second interpretation, it is *not* evident why the Dilemma Defender should assume that she is entitled to claim that the first horn is question begging.

II. TWO INTERPRETATIONS, TWO AUDIENCES, AND TWO DIALECTICAL SITUATIONS

The narrow interpretation of the Dilemma Defender's first horn is that, in those cases where the prior-sign assumes a deterministic relation, Frankfurt examples beg the question against the incompatibilist *because the deterministic relation expunges alternative possibilities.* It does so regardless of the other factors figuring in a Frankfurt example (factors such as counterfactual interveners). In contrast, on the broad interpretation, in those cases where the prior-sign assumes a deterministic relation, Frankfurt examples beg the question against the incompatibilist *because a deterministic relation obtains—simpliciter.* Understood this way, the concern is that, *independently of the importance of alternative possibilities,* Frankfurt-type examples that assume a deterministic relation cannot be used to convince *incompatibilists* that free will and moral responsibility do not require alternative possibilities. According to the incompatibilist, a *determined* agent is not free or morally responsible in the first place.

We now elaborate these interpretations. Beginning with the broad interpretation, suppose that, dialectically, the situation is this: the Frankfurt Defender means to persuade incompatibilists, who are already theoretically *committed* to a position regarding the free will problem.[12] In this dialectical situation, all that the Frankfurt Defender wants the *incompatibilist* to accept is that, if determinism *does* rule out the freedom required for moral responsibility, it is *not* by virtue of ruling out alternative possibilities. So the Frankfurt Defender reasons with the incompatibilist as follows:

---

[12] We ignore here the yet *further* dialectical context in which the Frankfurt defender means to persuade the committed compatibilist. This context has its own unique problems, but it does not center on contesting the assumption that an agent in a Frankfurt example is determined. For the best compatibilist resistance to Frankfurt examples, see Bernard Berofsky's "Classical Compatibilism. Not Dead Yet," in Widerker and McKenna, eds., pp 107–26.

> Let us work with whatever brand of libertarianism you want (noncausal, agent causal, or even indeterministic). With one exception, I will grant all its constituents to you, *including the formal incompatibilist requirement of indeterminism*. The exception is just the very point I want to contest—the necessity of alternative possibilities. On *that*, I request that you suspend judgment. Now let me introduce an example that should convince you that alternative possibilities are irrelevant to moral responsibility and the sort of freedom required for it.

Call this the *Broad Dialectical Situation*. On this interpretation of the situation, it should be clear that the Dilemma Defender's appeal to the first horn *is unimpeachable*; it is simply a condition of the debate that the Frankfurt Defender offer a case that does not presume determinism.

Why not couple a committed incompatibilist audience with a narrow interpretation of the first horn? Assuming that Frankfurt's argument is directed to a committed incompatibilist audience, it would be dialectically unfitting to assume that the first horn be assigned a narrow interpretation. If the Frankfurt Defender is going to grant libertarian freedom to an incompatibilist audience, and if the incompatibilist is going to allow Frankfurt a fair hearing, then the incompatibilist must be willing to put something on the table. She must be willing to concede that there is a subject of dispute which she is not simply entitled to treat in incompatibilist terms. If, in this dialectical context, what she would find objectionable about a Frankfurt example is the very fact that an agent in it had no alternative possibilities, then she would not in fact be prepared to suspend judgment on the very point Frankfurt wishes to dispute.

A further observation about the Broad Dialectical Situation merits attention. Suppose that the *compatibilist* Frankfurt Defender succeeds in convincing the incompatibilist that free and morally responsible agency does not require alternative possibilities.[13] The compatibilist Frankfurt Defender would *then* want to advance *further* arguments in opposition to the incompatibilist. This is because the sort of incompatibilist prepared to consider Frankfurt examples in this dialectical context would *not* be willing to concede that if Frankfurt's argument is sound, incompatibilism is *therefore* false.[14] Hence, in the Broad Dialectical Situation, this sort of incompatibilist, to begin with, approaches the debate over the Frankfurt-type examples under the assumption that perhaps determinism could be shown to rule out freedom and moral responsibility *independently* of ruling out alternative possibilities. And while this sort of incompatibilist might have initially assumed, prior to entertaining Frankfurt's argument, that it is sufficient to argue for incompatibilism by arguing that determinism rules out alternative possibilities, she would not be willing to grant that it is necessary. Were she to think it *necessary* to argue for incompatibilism by fixing upon alternative possibilities, then it would not make sense for her to suspend her commitment to alternative possibilities while considering a case of a libertarian free agent acting in a Frankfurt example.

---

[13] In discussing the Broad Dialectical Situation, we are fixing upon the *compatibilist* Frankfurt Defender. This, however, is only to bring into relief the further debates that might arise between compatibilists and incompatibilists if Frankfurt's argument is successful. We recognize that Frankfurt's argument is also available to incompatibilists who do not think that the incompatibility of determinism and free will turns upon alternative possibilities.

Now consider a narrow interpretation of the first horn. As we argued above, the narrow interpretation of the first horn is not dialectically suited for a context in which Frankfurt's argument is meant to engage directly a committed incompatibilist audience. We therefore propose that the narrow interpretation is better suited for a context in which Frankfurt's argument is directed at those who are *undecided* about the free will problem, those who have not yet committed to compatibilism or incompatibilism. Indeed, irrespective of the compatibility question, this undecided audience is not even prepared to specify what sorts of constituents are necessary for the freedom moral responsibility requires. Hence, they have no settled convictions as to whether, for example, freely willed conduct requires higher-order identification with one's own springs of action, or if it requires responsiveness to reason, or if it requires alternative possibilities, and so forth. This audience amounts to what in the political landscape might be called "undecided voters."

Clearly, given *this* dialectical target—the undecided—a critic of Frankfurt's argument is *certainly* not entitled to the broad interpretation of the first horn. That is, the Dilemma Defender is not entitled to the complaint that the prior-sign in a Frankfurt example assumes a deterministic relation *simpliciter*. In this dialectical context, it is not to be assumed that this is objectionable; it might not be. On the other hand, relying upon the narrow interpretation, the Dilemma Defender might think it reasonable to object that if determinism already rules out alternative possibilities, then there is no work for a Frankfurt example to do. The counterfactual intervener adds nothing; the argument evaporates. We shall now develop this point.

---

[14] In fact, several prominent incompatibilists are in precisely this position. See, for example, Hunt; Pereboom, "Determinism *al Dente*," *Noûs*, xxix (1995): 21–45, and *Living without Free Will*; Stump, "Libertarian Freedom and the Principle of Alternative Possibilities"; and Linda Zagzebski, "Does Libertarian Freedom Require Alternative Possibilities?" *Philosophical Perspectives*, xiv (2000): 231–48

Given a narrow interpretation of the first horn, consider the following dialectical context: the Frankfurt Defender invites an undecided audience to imagine a case of an agent who, prima facie, performs an uncontroversially blameworthy action, an action for which she is, prima facie, morally responsible. This audience makes no theoretical demands upon the metaphysical nature of this agent but simply assumes for argument's sake that we have a normally functioning person. Furthermore, no assumption is made as to whether the agent is determined, whether compatibilism or incompatibilism is true, or anything of the sort. In fact, it remains open whether the thesis of hard determinism is true and, hence, whether any agent acts of her own free will or is morally responsible for anything. (In this case, agents would merely be prima facie morally responsible for their *seemingly* blameworthy actions.) The Frankfurt Defender then asks this audience to accept that an agent's freedom and responsibility, if indeed any agent possesses these things, does not require alternative possibilities. So the Frankfurt Defender adds a counterfactual intervener to the case to do the work of eliminating the disputed alternatives, inviting the undecided audience to see that the lack of alternatives is *thereby* shown to be irrelevant to the blameworthiness of an agent's conduct. Now, given this dialectical context, the Dilemma Defender might protest:

> In a case in which the agent is determined, on one (admittedly disputed) view, the elimination of alternatives is rendered complete prior to the addition of the counterfactual machinery. And if an undecided audience grants that such an agent is even prima facie free and morally responsible without that machinery, as the incompatibilists might see it, this audience will have been duped into granting Frankfurt's point at the outset. Hence, a condition of this debate is that, absent Frankfurt's counterfactual machinery, it must be beyond dispute that an agent does have alternative possibilities.

Call this the *Narrow Dialectical Situation*. What is crucial to this dialectical situation is that the disputed ground is over the reasoned judgment of the theoretically uncommitted. If the Frankfurt Defender can win their allegiance, then she can show that the traditional incompatibilists (and also the traditional compatibilists) have been wrong all along to presume that basic intuitions about free will and moral responsibility commandingly suggest that a freely willing agent must be able to do otherwise. Hence, the burden of proof would be shifted to those who asserted that it must.

In the Narrow Dialectical Situation, the Dilemma Defender's objection to the first horn is not without merit. However, it is unclear that it should amount to the charge that her opponent is begging the question. A Frankfurt Defender might insist that, dialectically, she is *not* inviting an undecided audience to imagine an agent who *does* act freely and is morally responsible, or who, absent the counterfactual intervener, *does* have alternative possibilities; she is only asking her undecided audience to acknowledge that these matters also require suspending judgment. Then the Frankfurt Defender can go to work on the Dilemma Defender's assumption that one manner of ruling out alternative possibilities (allegedly the assumption of determinism) renders other manners irrelevant to our understanding the conditions for free and morally responsible agency.

III. TWO TELLING RESPONSES TO THE DILEMMA DEFENSE

To appreciate the significance of the two dialectical situations, it will be helpful to consider two recent responses to the Dilemma Defense. Alfred Mele explains that, on a standard interpretation, the alternative possibilities at issue in PAP are "alternative *actions*, including intentional refrainings and instances of deciding and choosing."[15] Further, these possibilities are "actions like stealing from a poor-box and deciding not to steal from a poor-box, as opposed to more refined actions like '*S*'s stealing from a poor-box without having been made to do so by someone else' and '*S*'s deciding not to steal from a poor-box because Black made *S* decide not to do so'." Mele remarks:

> Now, if *PAP*...on the standard interpretation of alternative possibilities ...[is] falsified by some Frankfurt-style cases, then a thief...need not ever have been able to choose otherwise than he did at any point in his life, on the standard interpretation of choosing otherwise, in order to be morally responsible for his present choice and action. But notice that this is not to say that responsibility is compatible with *determinism*. For a traditional incompatibilist about determinism and moral responsibility, an agent's being deterministically caused to *A* suffices for his not being morally responsible for *A*-ing. So if Frankfurt-style cases are to persuade traditional incompatibilists that agents can be morally responsible for stealing from a poor-box, say, even though they could not have done otherwise, it must be a feature of these examples that the agents' stealing

---

[15] "Ultimate Responsibility and Dumb Luck," *Social Philosophy and Policy*, XVI (1999). 274–93, here p. 283.

> was not deterministically caused. In principle, a libertarian's incompatibilism might be motivated, not by the thought that determinism precludes our ever having been able to do otherwise than we did, but instead by the thought that in a deterministic world our actions...are ultimately causally ensured *consequences* of the laws of nature and states that obtained long before we were born (*ibid.*, p. 284).

The phrase in the passage, "So if Frankfurt-style cases are to persuade traditional Incompatibilists," bears noting. The phrase, along with Mele's insightful comments, provides convincing support that he means to be operating within the Broad Dialectical Situation.

In a co-authored piece with David Robb,[16] Mele's intriguing response to the Dilemma Defense modifies the Frankfurt-type example so that the fail-safe condition is not dependent upon any prior-sign. The condition triggers the desired action only in the counterfactual event that the agent does not freely perform the action. In fact, though, the agent performs the mental action of making a decision on her own without its being causally determined. The example includes two independent processes, each aimed at the same outcome, the agent's making some decision, such as the decision to rob a poor-box. One process is deterministic and will bring about the outcome under all circumstances save the one in which the other process brings about the outcome. The second is indeterministic—the agent's allegedly freely making the decision to rob the poor-box. As things actually unfold, the agent makes the decision herself. But if the indeterministic process were to fail to result in this outcome, the deterministic process would ensure that it occurs. The indeterministic process trumps the deterministic process *at* the moment of action. The agent's action is assumed to be a libertarian free one, but given the deterministic causal process that is trumped, the agent has no alternatives.

For our concerns, we note that it is partly but crucially in virtue of presupposing the Broad Dialectical Situation that the Mele/Robb response seems apt. (Whether it is convincing is another matter.[17]) Mele and Robb's modified no-prior-sign Frankfurt-type example is offered in the spirit of respecting their presumed interlocutor's concern that determinism may be inimical to free action and responsibility independently of expunging alternative possibilities. Hence, their approach makes no effort to challenge the first horn. As the Broad

---

[16] See Mele and Robb, "Rescuing Frankfurt-Style Cases," and also "BBs, Magnets, and Seesaws: The Metaphysics of Frankfurt-Style Cases."

[17] See, for example, Ginet's 2001 addendum to his "In Defense of the Principle of Alternative Possibilities," pp. 85–87

Dialectical Situation dictates, they take the Dilemma Defender's first horn as unimpeachable.

Consider, in contrast, a strikingly different approach to the first horn. John Martin Fischer writes that it would not be "dialectically kosher" to assume that a determined agent in a Frankfurt example is morally responsible.[18] However, Fischer requests that it remain an open question. Fischer's more modest suggestion is merely that the presence of Frankfurt's intervener would not by itself be sufficient to show that the agent is *not* morally responsible. Hence, if the agent in a Frankfurt example is not morally responsible, it is not due to the agent's having alternative possibilities ruled out *by way of the counterfactual intervener*. This, as Fischer cautiously put it, "*suggests* that if Jones [an agent in a Frankfurt example] had no alternative possibilities at all [due to the truth of causal determinism], this would be irrelevant to the grounding of his moral responsibility."[19] So Fischer's response to the first horn involves three steps. First, he acknowledges that the Frankfurt Defender is not entitled to the assumption that a determined agent in a Frankfurt example is free and morally responsible. Second, he calls attention to the irrelevance of the counterfactual apparatus. Its presence alone would not explain why an agent in a Frankfurt example is not responsible (if she is not). Third, he holds that the irrelevance of eliminating alternative possibilities by way of the counterfactual intervener merely *suggests* that other means of ruling out alternative possibilities are also irrelevant.

Suppose that the Dilemma Defender presses Fischer by insisting that, in a deterministic scenario, he (Fischer) has not imagined a case in which it is clearly the counterfactual intervener that is ruling out the agent's alternative possibilities. If the agent is determined, then, the Dilemma Defender might object, *that* is why the agent does not possess alternative possibilities. And it would not be dialectically kosher to conclude that alternative possibilities are irrelevant in this context.[20] Hence, to insure that an important question is not begged, the Dilemma Defender might continue, we need a case in which the *only* manner of ruling out alternative possibilities is by way of the counterfactual intervener.

We acknowledge that the Dilemma Defender has a point. The Frankfurt Defender's case would be more convincing—there would be less to contest—if it were beyond dispute that an agent's alternatives were ruled out *only* by virtue of the counterfactual intervener and not also by virtue of the truth of determinism. But is the Dilemma Defender entitled to claim that, in a case that does assume determinism, the Frankfurt Defender's argument rises to the level of begging the question? We believe that this turns upon the following assumption of the Dilemma Defender: if determinism makes it the case that an agent in a Frankfurt example cannot do otherwise, then the presence of the counterfactual intervener cannot aid in understanding the disputed condition for moral responsibility. Although we do not wish to evaluate the veracity of this assumption here, we agree with Fischer that it does not seem unreasonable—at least it is not question begging—for the Frankfurt Defender to reply to it as follows: the counterfactual intervener's presence in a deterministic scenario, while perhaps not necessary for ruling out alternative possibilities, is sufficient for doing so. Furthermore, its presence *is* irrelevant to any judgment that the agent is not free or morally responsible.[21] So why can this not help us to think about other factors that are alleged to rule out an agent's freedom or responsibility?[22]

---

[18] "Free Will and Moral Responsibility," *Journal of Ethics* (forthcoming). See also "Recent Work on Moral Responsibility," and "Frankfurt-Type Examples and Semi-Compatibilism."

[19] "Free Will and Moral Responsibility."

[20] Stewart Goetz has pressed Fischer in just this way—see his "Frankfurt-Style Counterexamples and Begging the Question;" *Midwest Studies in Philosophy*, xxix (forthcoming 2005). See Fischer's reply in "Free Will and Moral Responsibility."

Again, for our interests, we emphasize that Fischer's response to the first horn of the Dilemma Defense is designed to address a Narrow

---

[21] Some might insist that the presumption of the (alleged) irrelevance of the intervener in the case of a determined agent is weaker than in the case of the undetermined agent. In the case of the undetermined agent, assuming the intervener does rule out all robust alternatives, there will, minimally, be nonrobust alternatives (or morally insignificant alternatives), and these might be entailed by other conditions that are significant and necessary for moral responsibility (such as ultimacy). But this would be to shift the focus of the Frankfurt debate as most understand it. For a discussion of Frankfurt's argument in light of nonrobust alternatives, see Haji, "Flickers of Freedom, Obligation, and Responsibility," *American Philosophical Quarterly*, XL (2003): 287–302.

[22] In light of this discussion, consider Fischer's remarks:

> In the Frankfurt-type scenario [assuming a determined agent], *two* causes make it the case that Jones [the agent] is unable to choose otherwise at [a certain time] *T2*: the prior condition of world (together with the laws of nature) and Black's counterfactual intervention. What the examples show is that *the mere fact* that Jones is unable to choose otherwise does not *in itself* establish that Jones is not morally responsible for his choice. This is because Black's counterfactual intervention is one of the things that make it the case that Jones is unable to choose otherwise at *T2*, and yet it is irrelevant to the grounding of Jones's moral responsibility. Considering this factor (the counterfactual intervention), and bracketing any other factor that might make it the case that Jones is unable to choose otherwise at *T2*, it seems to me that Jones may well be morally responsible for his action. The mere fact that he lacks alternative possibilities, then, cannot in itself be the reason that Jones is not morally responsible, if indeed he is not morally responsible—"Free Will and Moral Responsibility," Section IV.5.

Dialectical Situation. His response cannot plausibly be aimed at a direct attempt to convince the committed incompatibilist to give up the demand for alternative possibilities. It would not be reasonable to ask this audience to consider a *determined* agent who is a candidate for free and morally responsible agency. Rather, Fischer's response is meant to sway those who have not settled on whether determinism rules free and morally responsible agency. Furthermore, his response clearly presupposes a narrow interpretation of the first horn. He means to show that one manner of ruling out alternative possibilities does not impugn a philosophical lesson we can draw from another manner of ruling them out.

IV. DIALECTICAL DELICACIES AND THE CLAIM TO CONTESTED GROUND

We have been exploring hitherto unnoticed dialectical delicacies within the debate over the status of Frankfurt's argument against PAP.[23] Our discussion has helped distinguish two very different dialectical situations, complete with distinct argumentative burdens and ground rules as to what can be assumed either in advancing Frankfurt's argument or, instead, in objecting to it.

Under the assumption that Frankfurt's argument is meant to address directly a committed incompatibilist audience, it *is* dialectically required that the Frankfurt Defender proceed by offering an example that does not assume that the agent in it is determined. This is a burden that the Frankfurt Defender must accept in this dialectical context. But in this context, the incompatibilist who is prepared to entertain Frankfurt's argument is not entitled to demand that (absent Frankfurt's counterfactual machinery) the relevant case involves an agent with alternative possibilities. Demanding this is tantamount to refusing Frankfurt a fair hearing. Hence, fitting Frankfurt's argument for a committed incompatibilist audience requires a Broad Dialectical Situation. In the Broad Dialectical Situation, what the Dilemma Defender finds objectionable about the deterministic horn of the posed dilemma is that determinism obtains simpliciter. Given the dialectical terms in which the Frankfurt Defender enters into *this* debate, we agree that the Dilemma Defender's claim to the deterministic horn is beyond dispute. It simply makes no sense for the Frankfurt Defender to consider resistance on this front. Hence, in this context, the Frankfurt Defender's only plausible option is to resist the Dilemma Defense on its other horn, the horn in which it is assumed that the agent is not determined. (This is precisely the strategy adopted by Mele.)

---

[23] To our knowledge, the only other discussion of varying dialectical assumptions is in Pereboom's recent article, "Source Incompatibilism and Alternative Possibilities," see pp. 190–93 Some of Pereboom's observations are similar to ours However, he offers a different diagnosis of the dialectical forces at work in the debate.

But we also wish to note that those incompatibilists who insist upon the Broad Dialectical Situation—requiring that Frankfurt begin with a case of a libertarian free agent—are dialectically committed at least to the possibility that determinism might rule out free will and moral responsibility for reasons other than ruling out alternative possibilities. If so, certain sorts of incompatibilists are not entitled to argue with Frankfurt within the parameters of the Broad Dialectical Situation. Some incompatibilists hold that arguments for incompatibilism *require* a defense of the incompatibility of determinism and alternative possibilities. They believe that the only reason that determinism rules out free will and moral responsibility is because it rules out alternative possibilities. If these incompatibilists wish to grapple with Frankfurt given a Broad Dialectical Situation, we believe that they do so in bad faith. By demanding that Frankfurt begin with a case of a libertarian free agent, they are demanding that Frankfurt provide a case in which any freely willing and morally responsible agent does have alternative possibilities. This of course is just what Frankfurt wants to disprove.

Hence, we challenge incompatibilists to consider their reasons for incompatibilism. If their only reasons have to do with alternative possibilities, they must come clean. In doing so we believe that they must concede that they are not entitled to the Dilemma Defense if it is meant to address a Broad Dialectical Situation. If these incompatibilists wish to rely upon the Dilemma Defense, then they must accept the conditions of the debate established by the Narrow Dialectical Context.[34]

---

[34] This challenge is especially poignant for Kane. He is hard to fit within the dialectical parties as we have sorted them out. On the one hand, Kane maintains that an alternative possibilities condition, an AP condition (a PAP-like condition), is only one of two conditions necessary for libertarian free will and moral responsibility. According to Kane, another is an ultimacy condition. Ultimacy requires that an agent's freely willed actions ultimately trace back to factors that originate in the agent and cannot be traced back to causally sufficient conditions obtaining independently of the agent (*The Significance of Free Will*, pp. 58–59). Hence, it seems that Kane is a theorist who can insist upon the Broad Dialectical Situation. He can argue that there are grounds—the grounds of ultimacy—independent from considerations of alternative possibilities that could be used to establish the incompatibility of free will and determinism. However, Kane argues that ultimacy itself requires dual control (control over different courses of action) and that dual control entails alternative possibilities—"The Dual Regress of Free Will and the Role of Alternative Possibilities," *Philosophical Perspectives*, XIV (2000): 57–79. On our view of the dialectical terrain, if Kane wishes to do battle with the Frankfurt Defender within the Broad Dialectical Context, then he should reconsider his claim that ultimacy entails alternatives. If he is unwilling to concede this point, we believe that he is forced to argue with the Frankfurt Defender in a Narrow Dialectical Context.

In the narrow context, the presumed target of Frankfurt's argument is the theoretically uncommitted voter. Thus, the Frankfurt Defender does not have to initiate the debate by beginning with the case of a libertarian free agent. In this context, suggesting that an agent in a Frankfurt example is determined (simpliciter) is not necessarily objectionable, and the Dilemma Defender is not entitled to this as a complaint. However, as we have argued, the Dilemma Defender might understandably object in this narrow context that a determined agent's alternatives are ruled out prior to the addition of Frankfurt's counterfactual machinery. So the Dilemma Defender *can* reasonably argue that the first horn of the dilemma objects to a determined agent *because* such agents (allegedly) lack alternative possibilities. However, we have also argued that the Dilemma Defender's objection, when understood in the context, is at least disputable. In this case, the Dilemma Defender does not have a clear claim to the charge that the Frankfurt Defender is begging the question. In this context, the Frankfurt Defender has a plausible reply: even if determinism rules out alternative possibilities, so does the presence of the Frankfurt's counterfactual machinery. The latter way of ruling out alternatives is irrelevant to an evaluation of an agent's freedom and responsibility, and this provides plausible grounds for the suggestion that the former is as well. Our concern here is not to settle this dispute, only to point out that, dialectically, in this context, the Dilemma Defender cannot claim clear victory with respect to the first horn. On this horn, in this context, her charge of begging the question is unfounded. There is a legitimate dispute here, and the Frankfurt Defender deserves a hearing. (This is precisely the strategy adopted by Fischer.)

So in one dialectical context, the Broad one, the Dilemma Defender is entitled to the complaint that the first horn of the dilemma is question begging. In another dialectical context, the Narrow one, the Dilemma Defender is still entitled to the complaint that on the first horn the agent is determined, but her complaint does not amount to the devastating charge that the Frankfurt Defender is begging the question. We have argued that only certain sorts of incompatibilists are entitled to the Broad Dialectical Context, namely those who acknowledge that determinism might undermine free will and determinism for reasons other than ruling out alternative possibilities. Incompatibilists unwilling to acknowledge this will have to operate within the Narrow Dialectical Situation, and at least with respect to the first horn of the Dilemma Defense, they have their work cut out for them. In this dialectical context, the Frankfurt Defender, such as Fischer, has a dialectically fair and reasonable reply.

V. CONCLUSION

In conclusion, differentiating between two dialectical situations, it is natural to reflect on the big question: What is the appropriate dialectical situation that the Frankfurt examples should be thought to address, the Broad or the Narrow? Of course, argumentative strategies can be put to different sorts of purposes, and we do not wish to suggest that there is one "correct" context that Frankfurt's argument should be taken to address. On the one hand, the Narrow Dialectical Situation will not likely impress a committed incompatibilist. The committed incompatibilist will likely object that an argument merely designed to win the hearts of the uncommitted voter cannot speak to a theorist with deeper libertarian commitments regarding the conditions of agency.[25] Hence, the Frankfurt Defender wishing to address directly this sort of incompatibilist should be prepared to work within the framework of the Broad Dialectical Situation. In this case, the Frankfurt Defender should accept that she will only be able to engage the incompatibilist if she begins a dialogue by doing battle on the libertarian's own turf. On the other hand, why should the disputed turf between compatibilists and incompatibilists be granted to incompatibilists in the first place—as if it is the compatibilists' burden to win it back? The Frankfurt Defender who initiates the dialectic within the Broad Dialectical Situation has already conceded that the compatibilists have the burden of proof—*they* need to overcome the intuitive presumption in favor of both the demand for alternative possibilities and an incompatibilist understanding of them. But this seems to initiate the debate in the wrong place. Frankfurt-type examples, it seems to us, are best understood as ones that question the pre-theoretical intuitions that inform one's theorizing about the issues of free action and moral responsibility. They are, in short, thought experiments to settle, first, what turf should be the subject of controversy between compatibilists and incompatibilists. *If* this is so, then the Narrow Dialectical Context is the appropriate one for understanding the force of Frankfurt's argument against the importance of alternative possibilities for free will and moral responsibility.

ISHTIYAQUE HAJI

University of Calgary

MICHAEL MCKENNA

Ithaca College

---

[25] For example, in *The Significance of Free Will*, Kane maintains the considerations of ultimacy are also needed for free and morally responsible agency and that ultimacy is incompatible with determinism.

COMMENTS AND CRITICISM

STOPPING TO REFLECT*

Our note is prompted by a recent article by Frank Arntzenius, "Some Problems for Conditionalization and Reflection."[1] Through a sequence of examples, that article purports to show limitations for a combination of two inductive principles that relate current and future rational degrees of belief: *Temporal Conditionalization* and *Reflection*:

(i) *Temporal Conditionalization* is the rule that, when a rational agent's corpus of knowledge changes between *now* and *later* solely by learning the (new) evidence, B, then coherent degrees of belief are updated using conditional probability according to the formula, for each event A,

$$P_{later}(A) = P_{later}(A \mid B) = P_{now}(A \mid B).$$

(ii) *Reflection*[2] between *now* and *later* is the rule that current conditional degrees of belief defer to future ones according to the formula that, for each event A,

$$P_{now}(A \mid P_{later}(A) = r) = r.^{3}$$

We will use the expression "Reflection holds with respect to the event A" to apply to this equality for a specific event A.

It is our view that neither of these principles is mandatory for a rational agent.[4] However, we do not agree with Arntzenius that, in the examples in his article, either of these two is subject to new

---

* Our research was carried out under NSF Grant DMS 0139911. We thank Joseph Halpern for alerting one of us (T.S.) to the Sleeping Beauty problem, independent of Frank Arntzenius's article.

[1] This JOURNAL, C, 7 (July 2003). 356–70.

[2] See Bas C. van Fraassen's "Belief and the Will," this JOURNAL, LXXXI, 5 (May 1984): 235–56. van Fraassen's *Reflection* has an antecedent in Michael Goldstein's "Prevision of a Prevision," *JASA*, LXXVIII (1983): 817–19.

[3] Here and through the rest of this note 'r' is a standard designator for a real number—this in order to avoid a Miller-styled problem—see David Miller's "A Paradox of Information," *British Journal for the Philosophy of Science*, XVII (1966): 59–61.

[4] We have argued, for example, that when (subjective) probability is finitely but not countably additive, then there are simple problems where (i) is reasonable, but where (i) precludes (ii). See our "Reasoning to a Foregone Conclusion," *JASA*, XCI (1996): 1228–36. Also, Isaac Levi argues successfully, we think, that (i) is not an unconditional requirement for a rational agent—see his "The Demons of Decision," *The Monist*, LXX (1987): 193–211.

0022-362X/04/0106/315–22

restrictions or limitations beyond what is already assumed as familiar in problems of stochastic prediction.

To the extent that a rational person does not know *now* exactly what she or he will know in the future, anticipating one's future beliefs involves predicting the outcome of a stochastic process. The literature on stochastic prediction relies on two additional assumptions regarding states of information and the temporal variables that index them[5]:

(iii) When $t_2 > t_1$ are two fixed times, then the information the agent has at t_2 includes all the information that she or he had at time t_1.[6] This is expressed mathematically by requiring that the collection of information sets at all times through the future form what is called a *filtration*.

Second, since the agent may not know *now* the precise time at which some specific information may become known in the future, then future times are treated as *stopping times*. That is:

(iv) For each time T (random or otherwise) when a prediction is to be made, the truth or falsity of the event $\{T \leq t\}$ is known at time t, for all fixed t. Such (random) times T are called *stopping times*.

In this note, we apply the following three results[7] to the examples in Arntzenius's article. These results, we believe, help to explain why the examples at first appear puzzling and why they do not challenge either Temporal Conditionalization or Reflection. Result 1 covers the ordinary case, where Reflection holds. Results 2 and 3 establish that Reflection will fail when one or the other of the two additional assumptions, (iii) and (iv), fail. It is not hard to locate where these assumptions are violated in the examples that Arntzenius presents.

Result 1. When "*later*" is a *stopping time*, when the information sets of future times form a *filtration*, and assuming that degrees of belief are updated by Temporal Conditionalization, then Reflection between *now* and *later* follows.

Result 2. When the information known to the agent over time *fails* to form a filtration, not only is Temporal Conditionalization vacuously satisfied (as its antecedent fails), but then Reflection fails unless what is forgotten in the failure of filtration becomes practically certain (its probability becomes 0 or 1) in time for future predictions, *later.*

---

[5] See, for example, Patrick Billingsley, *Probability and Measure*, 3rd edition (New York: Wiley, 1995), section 35.

[6] Here and through the rest of this note, 't' is a standard designator for a real number for time. More precisely, we use the subscripted variable, for example 't_1' to denote a specific time as the agent of the problem is able to measure it. We presume that the agent has some real-valued "clock" that quantifies a transitive relation of "is later than." Subtleties about the differences between how time is so indexed for different observers is relevant to one of Arntzenius's puzzles, to wit, the Prisoner's Problem.

[7] Proofs for these three are given in the appendix. In this note, we assume that all probability is countably additive.

Result 3. However, if the information known to the agent over time forms a filtration and Temporal Conditionalization holds, but "*later*" is not a *stopping time,* then Reflection between *now* and *later* holds for the specific event A, that is, $P_{now}(A \mid P_{later}(A) = r) = r$, subject to the necessary and sufficient condition, (3.1), below.

Let H_t be the event "t = *later.*" When *later* is not a stopping time, the event H_t is news to the agent making the forecasts. The question at hand is whether this news is relevant to the forecasts expressed by Reflection. To answer that question, concerning such forecasts about the event A, define the quantity $y_t(A)$ by

$$y_t(A) = \frac{P_{now}(H_t \mid P_t(A) = r \,\&\, A)}{P_{now}(H_t \mid P_t(A) = r)}.$$

The quantity $y_t(A)$ is an index of the current conditional dependence between A and H_t, given that $P_t(A) = r$. For example, $y_t(A) = 1$ if and only if A and H_t are conditionally independent for the agent, *now,* given that $P_t(A) = r$. In other words, by symmetry of conditional independence, $y_t(A) = 1$ if and only if the agent's current conditional probability of A given that $P_{later}(A) = r$, is unchanged by the added information H_t.

Reflection holds for A between *now* and *later,* $P_{now}(A \mid P_{later}(A) = r) = r$ if and only if, given $P_{later}(A) = r$, the conditional expected value $y_T(A) = 1$. Specifically, if and only if

$$(3.1)\quad 1 = \Sigma_t\, y_t(A)\, P_{now}(H_t \mid P_{later}(A) = r)$$

Thus, Reflection is satisfied between *now* and *later* if and only if (3.1) holds for each A.

Next, we illustrate the second and third results with examples that show how Reflection may fail.

Example 1 (illustrating Result 2). Suppose that the agent will observe a sequence of coin tosses, one at a time at a known rate, for example one toss per minute. Let $X_n = 1$ if the coin lands heads up on toss n, and let $X_n = 0$ otherwise. The agent does not know how the coin is loaded, but believes that it is *fair* (event A) with personal probability 1/2, and that with personal probability 1/2 it is biased with a chance of 3/4 for landing tails (event A^c). Also, he believes that tosses are conditionally independent given the loading, that is, given that the coin is *fair* or given that it is biased 3/4 for tails.

Time is indexed for the agent by the number of the most recent coin toss. The time "*now*" occurs after the first toss ($t = n = 1$), and "*later*" denotes the time ($t = n = 2$) just after the second toss. Unfortunately, at each time t, the agent knows that he can remember only the most recent flip, X_t, though he knows which numbered toss it is because, for example, he can see a clock. Suppose that the first toss lands heads up, which is the event $C = \{X_1 = 1\}$. The information that will be available to the forgetful agent later (at $t = 2$) will be only that either $B_1 = \{X_2 = 1\}$ or $B_0 = \{X_2 = 0\}$. He will not recall C because of his predictable memory lapse, and he knows all this. It is straightforward to compute:

$$P_{later}(A \mid B_1) = 2/3 \text{ and } P_{later}(A \mid B_0) = 2/5.$$

However, at $t = 1$, the agent's conditional probability for A, given event B_1 occurring at $t = 2$, satisfies $P_{now}(A \mid B_1) = 4/5$. Similarly, if *now* he conditions on event B_0 occurring at $t = 2$, his conditional probability will satisfy $P_{now}(A \mid B_0) = 4/7$.

Of course, Temporal Conditionalization holds vacuously at the *later* time, since the information sets available to the agent do not form a filtration. Reflection fails in this setting, as the agent does not remember at the *later* time what happened *now*, and he knows this all along. If B_1 occurs then $P_{later}(A) = P_{later}(A \mid B_1) = 2/3$, and if B_0 occurs then $P_{later}(A) = P_{later}(A \mid B_0) = 2/5$. Hence,

$$P_{now}(A \mid P_{later}(A) = 2/3) = 4/5$$

and

$$P_{now}(A \mid P_{later}(A) = 2/5) = 4/7.$$

Example 2 (illustrating Result 3 when condition (3.1) fails and then Reflection fails too). Modify Example 1 so that the agent has no memory failures and updates his degrees of belief by Temporal Conditionalization. Also, change the time "*now*" to denote the minute prior to the first toss, that is, *now* is $t = n = 0$. Define the time "*later*" to be the *random time, T, just prior to the first toss that lands heads up*. From the point of view of the agent, the quantity T is not an observable random variable up to and including time T, and it is not a *stopping time* either. It is observable to the agent starting with time $T+1$, of course, as by then he will have seen when the first head occurs.

With probability 1 the possible values for T are $T = 0, 1, 2, \ldots$. It is straightforward to verify that: $P_{later}(A) = [1 + (3/2)^n]^{-1}$, when $T = n$, for $n = 0, 1, 2, \ldots$. Notice that $P_{later}(A) \leq 1/2$, no matter when T occurs, and $P_{later}(A) < 1/2$ for $T > 0$, since if $T > 0$, the initial sequence of tosses that the agent observes all land tails up. However, from the value of $P_{\text{later}}(A)$ and knowing it is this quantity, one may calculate T exactly and thus know the outcome of the $n+1^{\text{st}}$ toss, which is heads. But when the agent computes $P_{\text{later}}(A)$ at the time *later*, he does not then know that *later* has arrived. Thus, *later*, he is not in a position to use the extra information that he would get from knowing when T occurs to learn the outcome of the $n+1^{\text{st}}$ toss. To repeat the central point, T is not a stopping variable.

It is evident that Reflection fails, $P_{\text{now}}(A \mid P_{\text{later}}(A) = r) \neq P_{\text{later}}(A)$. The extra information, namely that $P_{\text{later}}(A) = r$ rather than merely that $P_t(A) = r$ where t is the time on the agent's clock, is information that is relevant to his *current* probability of A, since it reveals the outcome of the next toss. Even *now*, prior to any coin tosses, when he computes $P_{\text{now}}(A \mid P_{\text{later}}(A) = r)$, the conditioning event reveals to him the value of T, since n is a function of r. In this case, the conditioning event entails the information of n and when the first heads occurs, namely, on the $n+1^{\text{st}}$ toss. Then Reflection fails as

$$P_{\text{now}}(A \mid P_{\text{later}}(A) = [1 + (3/2)^n]^{-1}) = (1 + 3^n/2^{n+1})^{-1}.$$

It remains only to see that (3.1) fails as well. Consider the quantity $y_t(A)$ used in condition (3.1). $y_t(A) = \dfrac{P_{\text{now}}(H_t \mid P_t(A) = r \,\&\, A)}{P_{\text{now}}(H_t \mid P_t(A) = r)}$. Given $P_t(A) = r$, the added information that A obtains is relevant to the agent's current probability when *later* occurs. Specifically, as $P_t(A) = [1 + (3/2)^n]^{-1}$ entails that $t = n$, $P_{\text{now}}(H_t \mid P_t(A) = [1 + (3/2)^n]^{-1}) = P_{\text{now}}(X_{t+1} = 1 \mid P_t(A) = [1 + (3/2)^n]^{-1}) = (1/2)[1+(3/2)^n]^{-1} + (1/4)(3/2)^n[1+(3/2)^n]^{-1} < 1/2 = P_{\text{now}}(X_{t+1} = 1 \mid P_t(A) = [1+(3/2)^n]^{-1} \,\&\, A) = P_{\text{now}}(H_t \mid P_t(A) = [1 + (3/2)^n]^{-1} \,\&\, A)$. Thus, $y_t > 1$. Hence, $1 < \Sigma_t\, y_t(A)\, P_{\text{now}}(H_t \mid P_{\text{later}}(A) = r)$.

Example 3 (illustrating Result 3 when (3.1) obtains and Reflection holds even though *later* is not a *stopping time*). In this example, consider a sequence of three times, $t = 0$, 1, and 2. *Now* is time $t = 0$. The available information increases with time, so that the information sets form a filtration, and the agent updates his degrees of belief by Temporal Conditionalization. Let the random time *later* be one of the two times $t = 1$, or $t = 2$, chosen at random, but which one is not revealed to the agent. Let the event H_i be that *later* $= i$, ($i = 1$, 2) and suppose that the occurrence of H_i (or its failure) while not known to the agent at any of the three times is independent of all else that the agent does know at all three times. In this case, for each event A (even for $A = H_i$) equation (3.1) is satisfied. That is, by the assumptions of the problem, either $y_t(A) = \dfrac{P_{\text{now}}(H_t \mid P_t(A) = r \,\&\, A)}{P_{\text{now}}(H_t \mid P_t(A) = r)} = 1$, or if $A = H_t$ then $y_t(A) = \frac{P_{now}(H_t \mid P_{later}(A) = r)}{P_{now}(H_t \mid P_t(A) = r)} = 1$. Thus, $P_{now}(A \mid P_{later}(A) = r) = r$. That is, even though *later* is not a stopping time, Reflection holds in this case since, given that $P_{later}(A) = r$, no new (relevant) evidence about A is conveyed through knowing that *later* has arrived, H_t.

We note that Result 2 applies to the Sleeping Beauty,[8] Shangri La, and Duplication examples of Arntzenius's article, where known failures of memory are explicit to the puzzles. Result 3 applies to explain the failure of Reflection in the two versions of the "Prisoner" example where the *local time* in the story, as measured by an ordinary clock (for example, "11:30 PM" in John Collins's example) is not a *stopping time* for the Prisoner.

It is our impression of Collins's Prisoner example that the reader is easily mistaken into thinking that the *local time*, as measured by an ordinary clock in the story, is a *stopping time* for all the characters in the story. Then Reflection holds for each of them, in accord with Result 1. In Collins's example, the *local time*, for example, 11:30 PM, is a *stopping time* for the Jailor (and also for the reader), but *not* for the Prisoner. For the Prisoner, time is measured by real-valued increments over the starting point, denoted by "now." Increments of *local time* are *stopping times* for the Prisoner. This is because the Prisoner does not know at the start of the story which of two *local times* equals his time *now*. For subsequent times, he does know how much *local time* has elapsed since *now*. But that information is not equivalent to knowing the *local time*. That difference in what is a *stopping time* for different characters is what makes this a clever puzzle, we think.

APPENDIX

Proof of Result 1.[9] Assume that when X is a random variable and C is an event, the agent's expected value $E_P(X)$ and conditional expected value $E_P(X|C)$ exist with respect to the probability P. Let A be an event and let $X = P(A|Y)$ be a random variable, a function of the random variable Y. Then, as a consequence of the law of total probability, with C also a function of Y,

---

[8] See also J.Y. Halpern's "Sleeping Beauty Reconsidered: Conditioning and Reflection in Asynchronous Systems," Department of Computer Science, Cornell University (September 2003). We agree with Halpern that, in our words, coherence of a sequence of previsions does not require that they will be *well calibrated*—in a frequency sense of "well calibrated." That is, we think it is reasonable for Sleeping Beauty to give a prevision of 1/2 to the event that the known fair coin landed heads on the flip in question, each time she is woken up. What complicates the analysis is that the repeated trials in Sleeping Beauty's game do not form an independent sequence, and her mandated forgetfulness precludes any "feedback" about the outcome of past previsions. When repeated trials are dependent and there is no learning about past previsions, coherent previsions may be very badly calibrated in the frequency sense. For other examples and related discussion of this point see, for example, Seidenfeld, "Calibration, Coherence, and Scoring Rules," *Philosophy of Science*, LII (1985): 274–94.

[9] van Fraassen, "Belief and the Problem of Ulysses and the Sirens," *Philosophical*

$$(1.1) \quad P(A|C) = E_P[X|C].$$

Assume that the agent's degrees of belief *now* include his *later* degrees of belief as objects of uncertainty. That is, future events such as "$P_{\text{later}}(A) = r$" and "$P_{\text{later}}(A|C) = q$" are proper subjects, *now*, of the agent's current degrees of belief. Suppose that, *now*, the agent anticipates using (i) Temporal Conditionalization in responding to the new evidence $Y = y$ that he knows he will learn at the stopping time, *later*. For example, Y might be the result of a meter reading made at the *later* time, with a sample space of m many possible values $Y = \{y_1, \ldots, y_m\}$. Thus, by (i), for whichever value y of Y that results,

$$(1.2) \quad P_{\text{later}}(A) = P_{\text{later}}(A \mid Y = y) = P_{\text{now}}(A \mid Y = y).$$

Then, by (i) and (1.1), for C also a function of Y, the agent *now* believes that

$$(1.3) \quad P_{\text{now}}(A \mid C) = E_{P_{\text{now}}}[P_{\text{later}}(A) \mid C].$$

Let C be the event, "$P_{\text{later}}(A) = r$," which we presume is a possible value for $P_{\text{later}}(A)$ from the agent's current point of view. (This C is a function of Y.) Then, because *later* is a stopping time,

$$(1.4) \quad P_{\text{now}}(A \mid P_{\text{later}}(A) = r) = E_{P_{\text{now}}}[P_{\text{later}}(A) \mid P_{\text{later}}(A) = r)].$$

As

$$(1.5) \quad E_{P_{\text{now}}}[P_{\text{later}}(A) \mid P_{\text{later}}(A) = r)] = r,$$

therefore

$$(1.6) \quad P_{\text{now}}(A \mid P_{\text{later}}(A) = r) = r,$$

that is, then Reflection holds as well.

Proof of Result 2. To show that Reflection fails, consider two times $t_1 < t_2$. Call an event *forgotten* if its truth or falsity is known at time t_1 but not at time t_2. From the assumption that these times do not form a filtration, let E be forgotten between t_1 and t_2 and allow that at t_1

Studies, LXXVII (1995): 7–37, argues (pp. 17–19) that Temporal Conditionalization implies *Reflection*. His argument (pp. 18–19) has an additional, tacit assumption that the time t at which conditioning applies for *Reflection* is a stopping time.

this is known to happen at t_2. Since $P_{t_1}(E) \in \{0,1\}$, conditioning will not change this value, that is,

$$(2.1) \quad P_{t_1}(E) = P_{t_1}(E \mid P_{t_2}(E) = r)$$

for a set of r-values of probability 1 under P_{t_1}. But, since it is known at t_1 that E will be forgotten at t_2, $P_{t_1}(0 < P_{t_2}(E) < 1) = 1$. Hence Reflection fails as $0 < r < 1$ in (2.1).

Proof of Result 3. Assume that the agent's information sets form a filtration over time and that Temporal Conditionalization holds between *now* and *later* but that *later* is not a stopping time for the agent. Let H_t be the event "*later* = t" for the specific time t. That is, assume that $0 < P_{later}(H_t) < 1$, when *later* occurs at t.

Later is the future time we will focus on in calculating whether Reflection holds, that is, we will inquire whether, for each event A, $P_{now}(A \mid P_{later}(A) = r) = r$, or not. We calculate as follows.

$$P_{now}(A \mid P_{later}(A) = r)$$

$$= \Sigma_t P_{now}(A \,\&\, H_t \mid P_{later}(A) = r)$$

by the law of total probability.

$$= \Sigma_t P_{now}(A \mid P_{later}(A) = r \,\&\, H_t)\, P_{now}(H_t \mid P_{later}(A) = r)$$

by the multiplication theorem

$$= \Sigma_t \frac{P_{now}(H_t \mid P_t(A) = r \,\&\, A)}{P_{now}(H_t \mid P_t(A) = r)} P_{now}(A \mid P_t(A) = r)\, P_{now}(H_t \mid P_{later}(A) = r)$$

by Bayes's theorem and the equivalence of

$(P_{later}(A) = r \,\&\, H_t)$ and $(P_t(A) = r \,\&\, H_t)$

$$= r\, \Sigma_t \frac{P_{now}(H_t \mid P_t(A) = r \,\&\, A)}{P_{now}(H_t \mid P_t(A) = r)} P_{now}(H_t \mid P_{later}(A) = r)$$

as $P_{now}(A \mid P_t(A) = r) = r$ by Result 1.

$$= r\, \Sigma_t\, y_t(A)\, P_{now}(H_t \mid P_{later}(A) = r).$$

by the definition of $y_t(A)$

Hence, $P_{now}(A \mid P_{later}(A) = r) = r$ if and only if $\Sigma_t\, y_t(A)\, P_{now}(H_t \mid P_{later}(A) = r) = 1$, which is condition (3.1).

M.J. SCHERVISH, T. SEIDENFELD, J.B. KADANE
Carnegie Mellon University

NEW BOOKS

ALLEN, BARRY: *Knowledge and Civilization.* Boulder, CO: Westview, 2004. x, 342 p. Cloth $95.00, paper $35.00.

ALLISON, HENRY E.: *Kant's Transcendental Idealism: An Interpretation and Defense,* Revised and Enlarged Edition. New Haven: Yale, 2004. xix, 537 p. Paper $25.00.

ALTIERI, CHARLES: *The Particulars of Rapture: An Aesthetics of the Affects.* Ithaca, NY: Cornell, 2003. x, 299 p. Cloth $49.95, paper $22.50.

AMERIKS, KARL: *Interpreting Kant's "Critiques."* New York: Oxford, 2003. vii, 351 p. Cloth $74.00.

ARONSON, RONALD: *Camus and Sartre: The Story of a Friendship and the Quarrel that Ended It.* Chicago: University Press, 2004. x, 291 p. Cloth $32.50.

AZZOUNI, JODY: *Deflating Existential Consequence: A Case for Nominalism.* New York: Oxford, 2004. viii, 241 p. Cloth $49.95.

BAKER, GORDON: *Wittgenstein's Method: Neglected Aspects.* Malden, MA: Blackwell, 2004. xi, 312 p. Cloth $59.95.

BANKS, ERIK C.: *Ernst Mach's World Elements: A Study in Natural Philosophy.* Norwell, MA: Kluwer, 2003. xii, 292 p. Cloth $127.00.

BAUHN, PER: *The Value of Courage.* Lund, Sweden: Nordic, 2003. 213 p. Cloth $60.00.

BAUM, ERIC B.: *What Is Thought?* Cambridge: MIT, 2004. xi, 478 p. Cloth $40.00.

BEISER, FREDERICK C.: *The Romantic Imperative: The Concept of Early German Romanticism.* Cambridge: Harvard, 2003. xiii, 243 p. Cloth $45.00.

BEK, LISE: *Reality in the Mirror of Art.* Gylling, Denmark: Aarhus, 2004. 572 p. Cloth $65.95.

BELL, LINDA A.: *Beyond the Margins: Reflections of a Feminist Philosopher.* Albany: SUNY, 2003. x, 245 p. Cloth $59.50, paper $17.95.

BERTRAM, CHRISTOPHER: *Rousseau and the Social Contract.* New York: Routledge, 2004. x, 214 p. Paper $15.95.

BLOOM, PAUL: *Descartes' Baby: How the Science of Child Development Explains What Makes Us Human.* New York: Perseus, 2004. xv, 271 p. Cloth $26.00.

BODAN, MARGARET A.: *The Creative Mind: Myths and Mechanisms,* Second Edition. New York: Routledge, 2004. xiii, 344 p. Paper $21.95.

BOVENS, LUC, and STEPHAN HARTMANN: *Bayesian Epistemology.* New York: Oxford, 2003. ix, 159 p. Cloth $60.00, paper $27.50.

BRANN, EVA, with PETER KALKAVAGE, and ERIC SALEM: *The Music of the Republic: Essays on Socrates' Conversations and Plato's Writings.* Philadelphia: Paul Dry, 2004. xiv, 378 p. Cloth $24.95.

BRINK, DAVID O.: *Perfectionism and the Common Good: Themes in the Philosophy of T.H. Green.* New York: Oxford, 2003. xiv, 139 p. Cloth $27.50.

BROWN, JESSICA: *Anti-individualism and Knowledge.* Cambridge: MIT, 2004. xiv, 339 p. Paper $25.00.

BUNGE, MARIO: *Emergence and Convergence: Qualitative Novelty and the Unity of Knowledge.* Toronto: University Press, 2003. xiii, 330 p. Cloth $75.00.

CAMPBELL, JOSEPH: *The Hero with a Thousand Faces,* Second Edition. Princeton: University Press, 2004. lxvi, 403 p. Cloth $29.95.

CAPALDI, NICHOLAS: *John Stuart Mill: A Biography.* New York: Cambridge, 2004. xx, 436 p. Cloth $40.00.

CARRUTHERS, PETER: *The Nature of the Mind: An Introduction.* New York: Routledge, 2004. xii, 308 p. Paper $24.95.

CAVELL, STANLEY: *Cities of Words: Pedagogical Letters on a Register of the Moral Life.* Cambridge: Harvard, 2004. xi, 458 p. Cloth $29.95.

CHEAH, PHENG: *Spectral Nationality: Passages of Freedom from Kant to Postcolonial Literatures of Liberation.* New York: Columbia, 2003. xvi, 408 p. Cloth $62.50, paper $22.50.

CHIHARA, CHARLES S.: *A Structural Account of Mathematics.* New York: Oxford, 2004. xiv, 380 p. Cloth $70.00.

CLARK, DAVID K.: *Empirical Realism: Meaning and the Generative Foundation of Morality.* Lanham, MD: Lexington, 2004. ix, 419 p. Cloth $75.00.

CLARK, DESMOND M.: *Descartes's Theory of Mind.* New York: Oxford, 2003. viii, 267 p. Cloth $49.95.

CLARKE, RANDOLPH: *Libertarian Accounts of Free Will.* New York: Oxford, 2003. xv, 244 p. Cloth $49.95.

COOPER, DAVID E.: *Meaning.* Ithaca, NY: McGill-Queen's, 2004. 152 p. Cloth $70.00, paper $22.95.

COOPER, GREGORY J.: *The Science of the Struggle for Existence: On the Foundations of Ecology.* New York: Cambridge, 2003. xv, 319 p. Cloth $60.00.

CRAPANZANO, VINCENT: *Imaginative Horizons: An Essay in Literary-Philosophical Anthropology.* Chicago: University Press, 2004. xiii, 260 p. Cloth $50.00, paper $19.00.

CURTLER, HUGH MERCER: *Ethical Argument: Critical Thinking in Ethics,* Second Edition. New York: Oxford, 2004. xx, 184 p. Paper $19.95.

DA COSTA, NEWTON C. A., and STEVEN FRENCH: *Science and Partial Truth: A Unitary Approach to Models and Scientific Reasoning.* New York: Oxford, 2003. 259 p. Cloth $49.95.

DAVIES, BRIAN: *An Introduction to the Philosophy of Religion,* Third Edition. New York: Oxford, 2004. ix, 333 p. Paper $24.95.

DAVIES, DAVID: *Art as Performance.* Malden, MA: Blackwell, 2004. xii, 278 p. Paper $29.95.

DAY, JERRY: *Voegelin, Schelling, and the Philosophy of Historical Existence.* Columbia: Missouri UP, 2003. xiv, 300 p. Cloth $49.95

DEELY, JOHN: *The Impact on Philosophy of Semiotics.* South Bend, IN: St Augustine's, 2003. viii, 267 p. Cloth $26.00.

DEWITT, RICHARD: *Worldviews: An Introduction to the History and Philosophy of Science.* Malden, MA: Blackwell, 2004. xii, 326 p. Cloth $64.95, paper $29.95.

DICKERSON, A.B.: *Kant on Representation and Objectivity.* New York: Cambridge, 2004. x, 217 p. Cloth $60.00.

DILLON, JOHN, and LLOYD P. GERSON: *Neoplatonic Philosophy: Introductory Readings.* Indianapolis: Hackett, 2004. xxiv, 373 p. Cloth $47.95, paper $18.95.

DOMBROWSKI, DANIEL A.: *Divine Beauty: The Aesthetics of Charles Hartshorne.* Nashville: Vanderbilt, 2004. Cloth $59.95.

DROEGE, PAULA: *Caging the Beast: A Theory of Sensory Consciousness* Philadelphia: John Benjamins, 2003. x, 183 p. Paper $54.95.

DUMMETT, MICHAEL: *Truth and the Past.* New York: Columbia, 2004. x, 122 p. Cloth $29.50.

EDMUNDSON, WILLIAM A.: *An Introduction to Rights.* New York: Cambridge, 2004. xv, 223 p. Cloth $55.00, paper $22.00.

ELDER, CRAWFORD L.: *Real Nature and Familiar Objects.* Cambridge: MIT, 2004. Cloth $32.00.

ELDRIDGE, RICHARD: *An Introduction to the Philosophy of Art.* New York: Cambridge, 2003. ix, 285 p. Cloth $65.00, paper $23.00.

ENÇ, BERENT: *How We Act: Causes, Reasons, and Intentions.* New York: Oxford, 2003. xiii, 252 p. Cloth $45.00.

FALK, ARTHUR E.: *Desire and Belief: Introduction to Some Recent Philosophical Debates.* Lanham, MD: UP of America, 2004. xxi, 595 p. Paper $62.00.

FINE, GAIL: *Plato on Knowledge and Forms: Selected Essays.* New York: Oxford, 2003. ix, 447 p. Cloth $99.00.

FINKELSTEIN, DAVID H.: *Expression and the Inner.* Cambridge: Harvard, 2003. ix, 182 p. Cloth $45.00.

FLEISCHACKER, SAMUEL: *On Adam Smith's Wealth of Nations: A Philosophical Companion.* Princeton: University Press, 2004. xvii, 329 p. Cloth $39.50.

FLEMING, BRUCE E.: *Art and Argument. What Words Can't Do and What They Can.* Lanham, MD: UP of America, 2003. vi, 201 p. Paper $34.00.

FODOR, JERRY A.: *Hume Variations.* New York: Oxford, 2003. 165 p. Cloth $22.00.

FOSTER, JOHN: *The Divine Lawmaker: Lectures on Induction, Laws of Nature, and the Existence of God.* New York: Oxford, 2004. viii, 191 p. Cloth $45.00

FRANKFURT, HARRY G.: *The Reasons of Love.* Princeton: University Press, 2004. 100 p. Cloth $19.95.

GANDER, ERIC M.: *On Our Minds: How Evolutionary Psychology Is Reshaping the Nature-versus-Nurture Debate.* Baltimore: Johns Hopkins, 2003. x, 293 p. Cloth $45.00.

GÄRDENFORS, PETER: *How Homo Became Sapiens: On the Evolution of Thinking.* New York: Oxford, 2003. ix, 240 p. Cloth $59.50.

GARVER, EUGENE: *For the Sake of Argument: Practical Reasoning, Character, and the Ethics of Belief.* Chicago: University Press, 2004. xi, 272 p. Paper $22.50.

GIER, NICHOLAS F.: *The Virtue of Nonviolence: From Gautama to Gandhi.* Albany: SUNY, 2004. xv, 224 p. Cloth $40.00.

GOLDFARB, WARREN: *Deductive Logic.* Indianapolis: Hackett, 2003. xv, 292 p. Cloth $32.50.

GOLOMB, JACOB: *Nietzsche and Zion.* Ithaca, NY: Cornell, 2004. xii, 274 p. Cloth $39.95.

GRANT, ROBERT: *Imagining the Real: Essays on Politics, Ideology, and Literature.* New York: Palgrave MacMillan, 2003. xvi, 248 p. Cloth $69.95.

GREEN, CELIA: *The Lost Cause: Causation and the Mind-Body Problem.* Cambridge, UK: Oxford Forum, 2003. xli, 241 p. Cloth.

GREEN, T.H.: *Prolegomena to Ethics.* DAVID O. BRINK, ed. New York: Oxford, 2003. cl, 477 p. Cloth $85.00, paper $29.95.

GIULIANI, MASSIMO: *A Centaur in Auschwitz: Reflections on Primo Levi's Thinking.* Lanham, MD: Lexington, 2003. ix, 111 p. Cloth $60.00, paper $24.95.

HALPERN, JOSEPH Y.: *Reasoning about Uncertainty.* Cambridge: MIT, 2003. xiv, 483 p. Cloth $50.00.

HALWANI, RAJA: *Virtuous Liaisons: Care, Love, Sex, and Virtue Ethics.* Chicago: Open Court, 2003. ix, 315 p. Paper $27.95.

HAMBLET, WENDY C.: *The Sacred Monstrous: A Reflection on Violence in Human Communities.* Lanham, MD: Lexington, 2004. xiv, 111 p. Cloth $65.00, paper $22.00.

HANNA, PATRICIA, and BERNARD HARRISON: *Word and World: Practice and the Foundations of Language.* New York: Cambridge, 2004. xii, 420 p. Cloth $85.00, paper $30.00.

HARRIS, LEE: *Civilization and Its Enemies: The Next Stage of History.* New York: Free Press, 2004. xix, 232 p. Cloth $26.00.

HASKINS, EKATERINA V.: *Logos and Power in Isocrates and Aristotle.* Columbia: South Carolina UP, 2004. xiii, 172 p. Cloth $34.95.

HAWTHORNE, JOHN: *Knowledge and Lotteries.* New York: Oxford, 2004. viii, 205 p. Cloth $35.00.

HERWITZ, DANIEL: *Race and Reconciliation.* Minneapolis: Minnesota UP, 2003. xvii, 227 p. Cloth $60.95, paper $21.95.

HETCHER, STEVEN A.: *Norms in a Wired World.* New York: Cambridge, 2004. xii, 420 p. Cloth $65.00.

HILLS, HELEN: *Invisible City: The Architecture of Devotion in Seventeenth-Century Neapolitan Convents.* New York: Oxford, 2004. xiii, 268 p. Cloth $60.00.

HIMMELFARB, GERTRUDE: *The New History and the Old: Critical Essays and Reappraisals,* Revised Edition. Cambridge: Harvard, 2004. 266 p. Paper $18.95.

HOY, DAVID COUZENS: *Critical Resistance: From Poststructuralism to Post-Critique.* Cambridge: MIT, 2004. viii, 274 p. Cloth $35.00.

HUGHES, CHRISTOPHER: *Kripke: Names, Necessity, and Identity.* New York: Oxford, 2004. xii, 247 p. Cloth $60.00.

JACKSON, FRANK, and PHILIP PETTIT, and MICHAEL SMITH: *Mind, Morality, and Explanation: Selected Collaborations.* New York: Oxford, 2004. xii, 427 p. Cloth $95.00.

JACOB, PIERRE, and MARC JEANNEROD: *Ways of Seeing: The Scope and Limits of Visual Cognition.* New York: Oxford, 2003. xix, 287 p. Cloth $124.50.

JACOBS, LESLEY A.: *Pursuing Equal Opportunities: The Theory and Practice of Egalitarian Justice.* New York: Cambridge, 2003. Cloth $65.00, paper $24.00.

JACQUETTE, DALE: *Pathways in Philosophy: An Introduction Guide with Readings.* New York: Oxford, 2003. xv, 555 p. Paper $56.95.

JAMES, WENDY: *The Ceremonial Animal: A New Portrait of Anthropology.* New York: Oxford, 2003. xxiii, 384 p. Cloth $45.00.

JANOWSKI, ZBIGNIEW: *Augustinian-Cartesian Index: Texts and Commentary.* South Bend, IN: St. Augustine's, 2004. xv, 275 p. Cloth $35.00.

JOHNSON, DAVID MARTEL: *How History Made the Mind: The Cultural Origins of Objective Thinking.* Chicago: Open Court, 2003 xi, 227 p. Paper $24.95.

KATZ, CLAIRE ELISE: *Levinas, Judaism, and the Feminine: The Silent Footsteps of Rebecca.* Bloomington: Indiana UP, 2003. xvi, 199 p. Cloth $49.95, paper $22.95.

KATZ, JERROLD J.: *Sense, Reference, and Philosophy.* New York: Oxford, 2004. xvi, 224 p. Cloth $49.95.

KEKES, JOHN: *The Illusions of Egalitarianism.* Ithaca. Cornell, 2003. xii, 228 p. Cloth $29.95.

KELLY, EUGENE: *The Basics of Western Philosophy.* Westport, CT: Greenwood, 2004. xii, 295 p. Cloth $75.00.

KOLNAI, AUREL: *On Disgust.* BARRY SMITH and CAROLYN KORSMEYER, eds. Chicago, Open Court, 2004. viii, 120 p. Paper $18.95.

KRAMER, MATTHEW H.: *The Quality of Freedom.* New York: Oxford, 2003. ix, 482 p. Cloth $45.00.

KUKATHAS, CHANDRAN: *The Liberal Archipelago: A Theory of Diversity and Freedom.* New York: Oxford, 2003. xii, 292 p. Cloth $45.00.

KÜNNE, WOLFGANG: *Conceptions of Truth.* New York: Oxford, 2003. xiii, 493 p. Cloth $74.00.

KUSPIT, DONALD: *The End of Art.* New York: Cambridge, 2004. xi, 208 p. Cloth $28.00.

LANDMAN, FRED: *Indefinites and the Type of Sets.* Malden, MA: Blackwell, 2004. xxvii, 269 p. Cloth $78.95, paper $43.95.

LAPORTE, JOSEPH: *Natural Kinds and Conceptual Change.* New York: Cambridge, 2004. x, 221 p. Cloth $70.00.

LAWLOR, LEONARD: *The Challenge of Bergsonism: Phenomenology, Ontology, Ethics.* New York: Continuum, 2003. xiii, 154 p. Cloth $105.00, paper $24.95.

LEAR, GABRIEL RICHARDSON: *Happy Lives and the Highest Good: An Essay on Aristotle's Nichomachean Ethics.* Princeton: University Press, 2004. ix, 238 p. Cloth $35.00.

LEWENS, TIM: *Organisms and Artifacts: Design in Nature and Elsewhere.* New York: Cambridge, 2004. xi, 183 p. Cloth $32.00.

LIU, SHU-HSIEN: *Essentials of Neo-Confucian Philosophy.* Westport, CT: Greenwood, 2003. xv, 167 p. Cloth $89.95.

LLEWELYN, JOHN: *Seeing through God: A Geophenomenology.* Bloomington: Indiana UP, 2004. xv, 193 p. Cloth $59.95, paper $24.95.

LLOYD, DAN: *Radiant Cool: A Novel Theory of Consciousness.* Cambridge: MIT, 2004. xvii, 357 p. Cloth $24.95.

LLOYD, G.E.R.: *Ancient Worlds, Modern Reflections: Philosophical Perspectives on Greek and Chinese Science and Culture.* New York: Oxford, 2004. xi, 222 p. Cloth $35.00.

LUCY, NIALL: *A Derrida Dictionary.* Malden, MA: Blackwell, 2004. xii, 183 p. Paper $24.95.

LUNTLEY, MICHAEL: *Wittgenstein: Meaning and Judgement.* Malden, MA: Blackwell, 2003. viii, 187 p. Paper $29.95.

MCFEE, GRAHAM: *Sport, Rules, and Values: Philosophical Investigations into the Nature of Sport.* New York: Routledge, 2004 x, 2002 p. Paper $24.99.

MCINERNY, RALPH: *Aquinas.* New York: Polity, 2004. vii, 160 p. Cloth $59.95, paper $24.95; *The Very Rich Hours of Jacques Maritain: A Spiritual Life.* Notre Dame: University Press, 2003. viii, 235 p. Cloth $32.00.

MELNYK, ANDREW: *A Physicalist Manifesto: Thoroughly Modern Materialism.* New York: Cambridge, 2003. xii, 328 p. Cloth $70.00.

MEYERS, DIANA TIETJENS: *Essays on Identity, Action, and Social Life.* Lanham, MD: Rowman and Littlefield, 2004. xix, 319 p. Cloth $75.00, paper $27.95.

MOLNAR, GEORGE: *Powers: A Study in Metaphysics.* STEPHEN MUMFORD, ed. New York: Oxford, 2003. xiv, 238 p Cloth $45.00.

MONTUSCHI, ELEONORA: *The Objects of Social Science.* New York: Continuum, 2003. ix, 160 p. Paper $29.95.

MORIARTY, MICHAEL: *Early Modern French Thought: The Age of Suspicion.* New York: Oxford, 2003. xii, 271 p. Cloth $75.00.

MORTON, ADAM: *Philosophy in Practice: An Introduction to the Main Questions,* Second Edition. Malden, MA. Blackwell, 2004. xv, 446 p. Paper $39.95.

NANCY, JEAN-LUC: *A Finite Thinking.* SIMON SPARKS, ed. Stanford: University Press, 2003. viii, 351 p. Cloth $60.00, paper $24.95.

NICHOLS, SHAUN, and STEPHEN P. STICH: *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds.* New York: Oxford, 2003. 237 p. Cloth $70.00, paper $24.95.

NOONAN, HARRY W.: *Personal Identity* New York: Routledge, 2003. xiii, 239 p. Paper $31.95.

NUSSBAUM, MARTHA C.: *Hiding from Humanity: Disgust, Shame, and the Law.* Princeton: University Press, 2004. xv, 413 p. Cloth $29.95.

OLIN, DORIS: *Paradox.* Ithaca: McGill-Queen's, 2003. x, 222 p. Cloth $70.00, paper $27.95.

O'MEARA, DOMINIC J.: *Platonopolis: Platonic Political Philosophy in Late Antiquity.* New York: Oxford, 2003. xi, 249 p. Cloth $55.00.

OTT, WALTER R.: *Locke's Philosophy of Language.* New York: Cambridge, 2004. vii, 159 p. Cloth $55.00.

PARKER, FRED: *Scepticism and Literature: An Essay on Pope, Hume, Sterne, and Johnson.* New York: Oxford, 2003. x, 290 p. Cloth $72.00.

PARTEE, BARBARA H.: *Compositionality in Formal Semantics: Selected Papers by Barbara H. Partee.* Malden, MA: Blackwell, 2004. xi, 331 p. Cloth $79.95, paper $44.95

PASKOW, ALAN: *The Paradoxes of Art: A Phenomenological Investigation.* New York: Cambridge, 2004. xi, 260 p. Cloth $65.00.

PEACOCKE, CHRISTOPHER: *The Realm of Reason.* New York: Oxford, 2004. x, 284 p. Cloth $35.00.

PIHLSTRÖM, SAMI: *Naturalizing the Transcendental: A Pragmatic View.* Amherst, NY: Humanity, 2003. 390 p. Cloth $49.00.

PLACE, U.T.: *Identifying the Mind: Selected Papers of U.T. Place.* GEORGE GRAHAM, and ELIZABETH R. VALENTINE, eds. New York: Oxford, 2004. xii, 215 p. Cloth $35.00.

POLGER, THOMAS W.: *Natural Minds.* Cambridge: MIT, 2004. xxvii, 294 p. Cloth $38.00.

POTTER, MICHAEL: *Set Theory and Its Philosophy: A Critical Introduction.* New York: Oxford, 2004. xiii, 345 p. Cloth $85.00, paper $29.95.

PRIOR, ARTHUR N.: *Papers on Time and Tense.* PER HASLE, PETER ØHRSTRØM, TORBEN BRAÜNER, and JACK COPELAND, eds. New York: Oxford, 2003. ix, 331 p. Cloth $72.00, paper $24.95.

PUTNAM, HILARY: *Ethics without Ontology.* Cambridge: Harvard, 2004. x, 160 p. Cloth $29.95.

PYLYSHYN, ZENON: *Seeing and Visualizing: It's Not What You Think.* Cambridge: MIT, 2003. xviii, 563 p. Cloth $50.00.

QUINE, W.V.: *Quintessence: Basic Readings from the Philosophy of W. V. Quine.* ROGER F. GIBSON, JR., ed. Cambridge: Harvard, 2004. x, 416 p. Cloth $39.95.

RAATZSCH, RICHARD: *Eigentlich Seltsames: Wittgensteins "Philosophische Untersuchungen": Band I: Einleitung und Kommentar PU 1-64.* Paderborn, Germany: Schöningh, 2003. 490 p. Cloth.

REGAN, TOM: *Empty Cages: Facing the Challenge of Animal Rights.* Lanham, MD: Rowman and Littlefield, 2004. xi, 229 p. Cloth $21.95.

ROBERT, JASON SCOTT: *Embryology, Epigenesis, and Evolution: Taking Development Seriously.* New York: Cambridge, 2004. xvi, 158 p. Cloth $60.00.

ROTHSTEIN, SUSAN: *Structuring Events: A Study in the Semantics of Lexical Aspect* Malden, MA: Blackwell, 2004. x, 206 p. Paper $43.95.

RUNDLE, BEDE: *Why There Is Something Rather than Nothing.* New York: Oxford, 2004. xii, 204 p. Cloth $45.00.

SAID, EDWARD W.: *Humanism and Democratic Criticism.* New York: Columbia, 2004. xvii, 154 p. Cloth $19.95.

SAMUELSON, NORBERT M.: *Jewish Philosophy: An Historical Introduction.* New York: Continuum, 2003. vii, 344 p. Cloth $125.00, paper $34.95.

SÁNCHEZ, GONZALO J., JR.: *Pity in Fin-de-Siècle French Culture: "Liberté, Egalité, Pitié."* Westport, CT: Praeger, 2004. x, 319 p. Cloth $67.95.

SANTAYANA, GEORGE: *The Letters of George Santayana: Book Five, 1933–1936.* WILLIAM G. HOLZBERGER, ed. Cambridge: MIT, 2003. lxxvii, 609 p. Cloth $60.00.

SCHAUER, FREDERICK: *Profiles, Probabilities, and Stereotypes.* Cambridge: Harvard, 2003. xiii, 359 p Cloth $29.95.

SCHIFFER, STEPHEN: *The Things We Mean.* New York: Oxford, 2003. 362 p. Cloth $74.00, paper $24.95.

SHAFER-LANDAU, RUSS: *Moral Realism: A Defense.* New York: Oxford, 2003. x, 322 p. Cloth $45.00.

SHAMDASANI, SONU: *Jung and the Making of Modern Psychology: The Dream of a Science.* New York: Cambridge, 2003. xvi, 387 p. Cloth $75.00, paper $28.00.

SHANKS, NIALL: *God, the Devil, and Darwin: A Critique of Intelligent Design Theory.* New York: Oxford, 2004. xiii, 273 p. Cloth $29.95.

SHAPIRO, LAWRENCE A.: *The Mind Incarnate.* Cambridge: MIT, 2004. xvi, 253 p. Cloth $34.00.

SHOEMAKER, SYDNEY: *Identity, Cause, and Mind,* Expanded Edition. New York: Oxford, 2003. x, 461 p. Paper $24.95.

SINGER, IRVING: *Three Philosophical Filmmakers: Hitchcock, Welles, Renoir.* Cambridge: MIT, 2004. x, 279 p. Cloth $32.95.

SOAMES, SCOTT: *Philosophical Analysis in the Twentieth Century,* Volume 2: *The Age of Meaning.* Princeton: University Press, 2003. xxii, 479 p. Cloth $35.00.

SOBEL, JORDAN HOWARD: *Logic and Theism: Arguments for and against Beliefs in God.* New York: Cambridge, 2004. xix, 652 p. Cloth $95.00.

SOLOMON, ROBERT C: *Living with Nietzsche: What the Great "Immoralist" Has to Teach Us.* New York: Oxford, 2003. x, 243 p. Cloth $35.00.

SOLSO, ROBERT L.: *The Psychology of Art and the Evolution of the Conscious Brain.* Cambridge, MIT, 2003. xvi, 278 p. Cloth $45.00.

SORENSEN, ROY: *A Brief History of the Paradox: Philosophy and the Labyrinths of the Mind.* New York: Oxford, 2003. xv, 394 p. Cloth $26.00.

SPITZER, MICHAEL: *Metaphor and Musical Thought.* Chicago: University Press, 2004. x, 380 p. Cloth $60.00.

STALEY, KENT W.: *The Evidence for the Top Quark: Objectivity and Bias in Collaborative Experimentation.* New York: Cambridge, 2004. xvi, 343 p. Cloth $70.00.

STALNAKER, ROBERT C.: *Ways a World Might Be: Metaphysical and Anti-Metaphysical Essays.* New York: Oxford, 2003. xii, 287 p. Cloth $72.00, paper $22.00.

STEWART, JON: *Kierkegaard's Relations to Hegel Reconsidered.* New York: Cambridge, 2003. xix, 695 p. Cloth $55.00.

STOUT, JEFFREY: *Democracy and Tradition.* Princeton: University Press, 2004. xv, 349 p. Cloth $35.00.

SULLIVAN, VICKIE B.: *Machiavelli, Hobbes, and the Formation of a Liberal Republicanism in England.* New York: Cambridge, 2004. x, 284 p. Cloth $75.00.

TALLIS, RAYMOND: *The Hand: A Philosophical Inquiry into Human Being.* Edinburgh: University Press, 2003. x, 364 p. Cloth $90.00, paper $30.00.

TAN, SOR-HOON: *Confucian Democracy: A Deweyan Reconstruction.* Albany: SUNY, 2003. viii, 258 p. Cloth $65.50, paper $20.95.

TAYLOR, CHARLES: *Modern Social Imaginaries.* Durham, NC: Duke, 2004. xi, 215 p. Paper $18.95.

TAYLOR, JOAN E.: *Jewish Women Philosophers of First-Century Alexandria: Philo's "Therapeutae" Reconsidered.* New York: Oxford, 2003. xv, 417 p. Cloth $99.00.

TAYLOR, PAUL C.: *Race: A Philosophical Introduction.* Malden, MA: Polity, 2004. xiv, 202 p. Cloth $62.95, paper $26.95.

TYE, MICHAEL: *Consciousness and Persons: Unity and Identity.* Cambridge: MIT, 2003. xv, 203 p. Cloth $35.00.

WARD, ROGER A.: *Conversion in American Philosophy: Exploring the Practice of Transformation.* New York: Fordham, 2004. xli, 235 p. Cloth $70.00.

WERTHEIMER, ALAN: *Consent to Sexual Relations.* xv, 293 p. New York: Cambridge, 2003. Cloth $70.00, paper $26.00.

WEST, HENRY R.: *An Introduction to Mill's Utilitarian Ethics.* New York: Cambridge, 2004. xi, 216 p. Cloth $55.00, paper $20.00.

WETTSTEIN, HOWARD: *The Magic Prism: An Essay in the Philosophy of Language.* New York: Oxford, 2004. x, 241 p. Cloth $49.95.

CAMBRIDGE

OUTSTANDING SCHOLARSHIP FROM CAMBRIDGE

The Philosophy of Biology
An Episodic History
Marjorie Grene and David Depew
The Evolution of Modern Philosophy
$75.00 Hardback 0-521-64371-6 440pp
$29.99 Paperback 0-521-64380-5

Isaiah Berlin, Letters 1928-1946
Edited by Henry Hardy
$40.00* Hardback
0-521-83368-X 752pp

Plato's Natural Philosophy
A Study of the *Timaeus-Critias*
Thomas Kjeller Johansen
$75.00 Hardback 0-521-79067-0 c 232pp

The Sociology of Norbert Elias
Edited by Steven Loyal and Stephen Quilley
$80.00 Hardback 0-521-82786-8 300pp
$29.00 Paperback 0-521-53509-3

Feelings and Emotions
The Amsterdam Symposium
Edited by Antony S.R. Manstead, Nico H. Frijda, and Agneta H. Fischer
Studies in Emotion and Social Interaction
$95.00 Hardback 0-521-81652-1 496pp
$35.00 Paperback 0-521-52101-7

The Cambridge Companion to Peirce
Edited by Cheryl Misak
Cambridge Companions to Philosophy
$70.00 Hardback 0-521-57006-9 376pp
$25.99 Paperback 0-521-57910-4

Autonomy
Edited by Ellen Frankel Paul, Fred D. Miller, Jr., and Jeffrey Paul
Social Philosophy and Policy
$25.00 Paperback 0-521-53499-2 358pp

Stoicism
Traditions and Transformations
Edited by Steven K. Strange and Jack Zupko
$70.00 Hardback 0-521-82709-4 312pp

Machiavelli, Hobbes, and the Formation of a Liberal Republicanism in England
Vickie B. Sullivan
$75.00 Hardback 0-521-83361-2 296pp

Spectacles of Truth in Classical Greek Philosophy
***Theoria* in its Cultural Context**
Andrea Wilson Nightingale
$70.00* Hardback
0-521-83825-8 c 300pp

Divine Motivation Theory
Linda Trinkaus Zagzebski
$70.00* Hardback
0-521-82880-5 c 300pp
$25.00* Paperback
0-521-53576-X

**Prices subject to change*

www.cambridge.org CAMBRIDGE UNIVERSITY PRESS

The
Southern Journal of
Philosophy

Spindel Conference 2004
Ancient Greek Ethics and Political Philosophy
Directed by Tim Roche
Supplement to Vol. XLIII (Spring 2005)

Kristeva's Social and Political Thought - Vol XLII
The Legacy of G. E. Moore: 100 Years of Metaethics - Vol XLI
Origins: The Common Sources of the Analytic and Phenomenological Traditions - Vol. XL
Hegel's *Philosophy of Right* - Vol. XXXIX
The Role of the Empirical (and of the A Priori) in Epistemology - Vol. XXXVIII
Nietzsche and Politics - Vol. XXXVII
Kant's *Metaphysics of Morals* - Vol. XXXVI
Rethinking Sex and Gender - Vol. XXXV
Explanation in the Human Sciences - Vol. XXXIV
Vagueness - Vol. XXXIII
Derrida's Interpretation of Husserl - Vol. XXXII
Ancient Minds - Vol. XXXI
Kant's Third Critique - Vol. XXX
Moral Epistemology - Vol. XXIX
Heidegger and Praxis - Vol. XXVIII
Aristotle's Ethics - Vol. XXVII
Connectionism - Vol. XXVI
B-Deduction - Vol. XXV
Moral Realism - Vol. XXIV
Recovering the Stoics - Vol. XXIII
Supervenience - Vol. XXII
Rationalist Conception of Consciousness - Vol. XXI

The Southern Journal of Philosophy
The University of Memphis
329 Clement Hall
Memphis TN 38152 USA
(901) 678-2669 Fax (901) 678-4365

http://cas.memphis.edu/philosophy/sjp

Ethics

an international journal of social political and legal philosophy

John Deigh, Editor

APA members save 20%

Purpose and Scope

Founded in 1890, ***Ethics*** is an international journal of moral, political, and legal philosophy. ***Ethics*** presents scholarly work from a variety of intellectual perspectives, including philosophy, social and political theory, law, and economics. The journal presents theory and its applications on contemporary moral issues. In addition, ***Ethics*** offers about 200 review essays, survey articles, book reviews, and book notes per volume.

Special Features

Ethics also holds symposia and publishes special issues that examine particular topics in depth through the contribution of distinguished scholars.

Ethics Electronic Edition

- Links to author references, tables, and figures
- Articles available for PDF printing capability
- Personalized passwords allow flexible access from any location

FREE Table of Contents Alerting

Sign up for a free table of contents alerting, please visit
www.journals.uchicago.edu/ET/journal/index.html

The University of Chicago Press

[illegible]

the review of

metaphysics

a philosophical quarterly

MARCH 2004 | **VOL. LVII, No. 3** | **ISSUE No. 227** | **$15.00**

articles

ARMAND MAURER	*Darwin, Thomists, and Secondary Causality*
C. V. MIRUS	*Aristotle's* agathon
LEO STRAUSS	*The Place of the Doctrine of Providence according to Maimonides*
ANTOON BRAECKMAN	*From the Work of Art to Absolute Realism: Schelling's Journey toward Absolute Idealism*
STEVEN B. SMITH	*An Exemplary Life: The Case of René Descartes*
ELIZABETH C. SHAW AND STAFF	*Summaries and Comments*

book reviews *philosophical abstracts* *announcements*

Individual Subscriptions $35.00 Institutional Subscriptions $55.00 Student/Retired Subscriptions $20.00

VISA AND MASTERCARD ACCEPTED

Jude P. Dougherty, Editor
The Catholic University of America, Washington D.C. 20064
email: mail@reviewofmetaphysics.org

Modern Philosophy

The Triumph of Venus

The Erotics of the Market

Jeanne Lorraine Schroeder

"A cornerstone to a whole new approach to legal theory." —Renata Salecl, author of *The Spoils of Freedom*

Philosophy, Social Theory, and the Rule of Law, $80.00 cloth

Thing Knowledge

A Philosophy of Scientific Instruments

Davis Baird

"Must reading for anyone interested in the development of science and its attendant technologies."

—Joseph C. Pitt, author of *Thinking About Technology*

$65.00 cloth

At bookstores or order (800) 822-6657 • www.ucpress.edu

UNIVERSITY OF CALIFORNIA PRESS

THE JOURNAL OF PHILOSOPHY

The Journal is pleased to announce to our readers and subscribers the launch of its new website, which will have the pages listed below. The address is: *www.journalofphilosophy.org* , or simply *www.jphil.org*.

- Purpose & History
- Contacts and Addresses
- Board of Trustees
- Editorial Board
- Memos to Authors
- Subscriptions & Advertising
- Recent & Forthcoming Issues
- Electronic Issues
- Index 1904–present
- Rights and Permissions
- Lectures & Conferences
- Philosophy Links

Prometheus Books

SIDNEY HOOK RECONSIDERED

Edited by Matthew J. Cotter

Foreword by Cornel West

Afterword by Richard Rorty

The essays in this volume are the outcome of a centennial celebration honoring the life and career of Sidney Hook. In addition to some of his former students, colleagues, allies and adversaries, this volume contains several essays by younger scholars, who offer fresh insights into Hook's philosophical significance.

This outstanding collection of essays presents an excellent reassessment of one of America's most misunderstood intellectuals.

400 pp • ISBN 1-59102-193-6 • HC $32

Humanity Books

an imprint of Prometheus Books

HABERMAS, NIETZSCHE, AND CRITICAL THEORY

Edited by Babette E. Babich

Critical theory has rejected challenges from postmodern thought, from Nietzscheans of every stripe, and from many of the social and political movements associated with alternate identities, such as feminism and gay-lesbian identity politics, among others.

The distinguished authors in this volume argue that critical theory is best served by an authentic engagement with these new developments and with Nietzsche. This important collection includes a newly revised version of Habermas's 1968 reflections on Nietzsche.

400 pp • ISBN 1-59102-209-6 • HC $79

HUSSERL ON ETHICS AND INTERSUBJECTIVITY

From Static to Genetic Phenomenology

Janet Donohoe

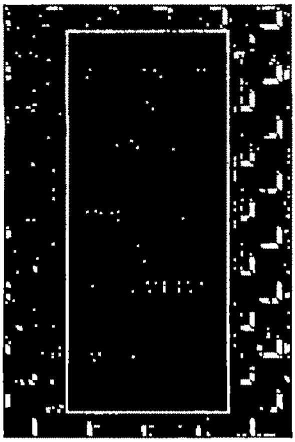
This book provides a compelling look at the importance of Husserl's methodological shift from his original, purely "static" approach to his later "genetic" approach to the analysis of consciousness.

Engaging critics from contemporary analytic schools to adherents of critical theory and deconstruction, to second and third generation phenomenologists, Donohoe shows that they often do not do justice to the breadth of Husserl's thought.

250 pp • ISBN 1-59102-210-X • HC $59

For exam copies contact LRisio@prometheusbooks.com
Call toll free (800) 421-0351 • Fax (716) 691-0137 • www.prometheusbooks.com

Incisive.



Wittgenstein on the Arbitrariness of Grammar

Michael N. Forster

What is the nature of a conceptual scheme? Are there alternative conceptual schemes? If so, are some more justifiable or correct than others? This book explores Wittgenstein's views on these questions.

Strongly argued and clearly written, this is a work that will appeal not only to philosophers but also to students of the human sciences, for whom Wittgenstein's work holds great relevance.

"Nuanced and convincingly supported, Forster's work reaches conclusions of great intrinsic interest."—Paul Horwich, University College, London and City University of New York

Cloth $29.95 ISBN 0-691-11366-1

The Humor of Kierkegaard

An Anthology

Søren Kierkegaard

Edited and Introduced by Thomas C. Oden

Who might reasonably be nominated as the funniest philosopher of all time? With this anthology, Thomas Oden provisionally declares Søren Aabye Kierkegaard—despite his enduring stereotype as the melancholy, despairing Dane—as the most amusing

This collection shows how Kierkegaard's theory of comedy is integrated into his practice of comic perception, and how both are integral to his entire authorship.

Paper $16.95 ISBN 0-691-02085-X
Cloth $55.00 ISBN 0-691-07406-2

PRINCETON
University Press

800-777-4726 • READ EXCERPTS ONLINE
WWW.PUP.PRINCETON.EDU

New from

Oxford

New in Paperback!

Thinking about Consciousness

DAVID PAPINEAU

"An eminently clear book."—*Choice*

The relation between subjective consciousness and the physical brain is widely regarded as the last mystery facing science. Papineau argues that consciousness seems mysterious not because of any hidden essence, but only because we think about it in a special way. He exposes the resulting potential for confusion, and shows that much scientific study of consciousness is misconceived.

2002 (paper 2004) paper $19.95 cloth $35.00

Three Faces of Desire

TIMOTHY SCHROEDER

Desires lead to actions, influence feelings, and determine what counts as a reward. Recent empirical evidence shows that these three aspects of desire stem from a common biological origin. Informed by contemporary science as much as by the philosophical tradition, *Three Faces of Desire* reveals this common foundation and builds a striking new philosophical theory of desire that puts desire's neglected face—reward—at its core.

(Philosophy of Mind Series)

2004 $35.00

Pyrrhonian Skepticism

Edited by WALTER SINNOTT-ARMSTRONG

Throughout the history of philosophy, skepticism has posed one of the central challenges of epistemology. This thematically unified volume presents previously unpublished articles on Pyrrhonian skepticism by a strikingly impressive group of philosophers, who engage with both historical and contemporary versions of the subject.

2004 $39.95

Descriptions and Beyond

Edited by MARGA REIMER and ANNE BEZUIDENHOUT

Marga Reimer and Anne Bezuidenhout present a collection of brand-new essays on important topics at the intersection of philosophy and linguistics. Written by a stellar line-up of contributors drawn from both disciplines, the papers will likewise attract a wide readership of professionals and students from either side.

2004 paper $35.00 cloth $99.00

Why There Is Something Rather than Nothing

BEDE RUNDLE

The question "Why is there something rather than nothing?" has a strong claim to be philosophy's central, and most perplexing, question. Bede Rundle challenges the stalemate between theistic and naturalistic explanations with a rigorous, properly philosophical approach, and presents some startlingly novel conclusions.

2004 $45.00

Prices are subject to change and apply only in the US. To order, or for more information, please call 1-800-451-7556. Visit our website at www.oup.com/us

OXFORD